A Comprehensive Review of the Load Forecasting Techniques Using Single and Hybrid Predictive Models

ABDULLAH AL MAMUN[ID][1], **MD. SOHEL**[ID][1], **NAEEM MOHAMMAD**[ID][2], **(Student Member, IEEE),**
MD. SAMIUL HAQUE SUNNY[ID][2], **DEBOPRIYA ROY DIPTA**[ID][2],
AND EKLAS HOSSAIN[ID][3], **(Senior Member, IEEE)**

[1]Department of Electrical and Electronics Engineering, International Islamic University Chittagong, Chattogram 4318, Bangladesh
[2]Department of Electrical and Electronics Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh
[3]Department of Electrical Engineering and Renewable Energy, Oregon Renewable Energy Center (OREC), Oregon Tech, Klamath Falls, OR 97601, USA

Corresponding author: Eklas Hossain (eklas.hossain@oit.edu)

ABSTRACT Load forecasting is a pivotal part of the power utility companies. To provide load-shedding free and uninterrupted power to the consumer, decision-makers in the utility sector must forecast the future demand for electricity with a minimum error percentage. Load prediction with less percentage of error can save millions of dollars to the utility companies. There are numerous Machine Learning (ML) techniques to amicably forecast electricity demand, among which the hybrid models show the best result. Two or more than two predictive models are amalgamated to design a hybrid model, each of which provides improved performances by the merit of individual algorithms. This paper reviews the current state-of-the-art of electric load forecasting technologies and presents recent works pertaining to the combination of different ML algorithms into two or more methods for the construction of hybrid models. A comprehensive study of each single and multiple load forecasting model is performed with an in-depth analysis of their advantages, disadvantages, and functions. A comparison between their performance in terms of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) values are developed with pertinent literature of several models to aid the researchers with the selection of suitable models for load prediction.

INDEX TERMS Load forecasting, predictive models, machine learning, support vector machines, artificial neural networks, computational intelligence, power industry, smart grid.

NOMENCLATURE
ABBREVIATIONS

ABC	Artificial Bee Colony
AIN	Artificial Immune Network
AIS	Artificial Immune System
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
ARMAX	Auto Regressive Moving Average with Exogenous Input
BA	Bat Algorithm
BFGS-FA	Broyden-Fletcher-Goldfarb-Shanno-Firefly Algorithm
BP	Back Propagation
CI	Computational Intelligence
CNN	Convolutional Neural Network
CT	Clustering Technique
DAE	Deep Auto Encoder
DAF	Dynamic Activation Function
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
D-RNN	Deep Recurrent Neural Network
EEMD	Ensemble Empirical Mode Decomposition
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
FA	Firefly Algorithm
FCM	Fuzzy C-Mean

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao[ID].

FCW	Fuzzy Combination Weight
FL	Fuzzy Logic
FOA	Fruitfly Optimization Algorithm
FTS	Fuzzy Time Series
FRBS	Fuzzy Rule-base System
GA	Genetic Algorithm
GAF	Genetic Algorithm with Fuzzy Logic
GHSA	Global Harmony Search Algorithm
GNN	Generalized Neural Network
GOA	Grasshopper Optimization Algorithm
GP	Genetic Programming
GRNN	Generalized Recurrent Neural Network
GRU	Gated Recurrent Unit
GSA	Gravitational Search Algorithm
HGASVR	Hybrid Genetic Based Support Vector Machine
HS	Harmony Search
IA	Immune Algorithm
IAGA	Improved Adaptive Genetic Algorithm
IEMD	Improved Empirical Mode Decomposition
IGSA	Improved Gravitational Search Algorithm
IMF	Intrinsic Mode Function
IS	Immune System
IT2FLS	Interval Type-2 Fuzzy Logic System
KF	Kalman Filter
LM	Levenberg-Marqardt
LTLF	Long-term Load Forecasting
LS	Least Squares
LSTM	Long Short-term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MFA	Modified Firefly Algorithm
MGGP	Multi-Gene Genetic Programming
ML	Machine Learning
MLP	Multilayer Perceptron
MSCNN	Multi Scale Convolutional Neural Network
MTLF	Medium-term Load Forecasting
NARX	Nonlinear Autoregressive Models with Exogenous Input
NFIS	Neural Fuzzy Inference System
NN	Neural Network
PDRNN	Pooling-based Deep Recurrent Neural Network
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
RELM	Recurrent Extreme Learning Machine
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RT	Regression Tree
SA	Stimulated Annealing
SAF	Static Activation Function
SARIMA	Seasonal Autoregressive Integrated Moving Average
SLFN	Single-hidden Layer Feed-forward Neural Network
SOM	Self-Organizing Map
STLF	Small Term Load Forecasting
SVM	Support Vector Machine
SVR	Support Vector Regression
THI	Temperature Humidity Index
TSK-FIS	Takagi-Sugeno-Kang Fuzzy Inference System
VaR	Value at Risk
VMD	Variational Mode Decomposition
VSTLF	Very Short-term Load Forecasting
WCI	Wind Chill Index
WNN	Wavelet Neural Network
WT	Wavelet Transform

SYMBOLS

y_i	Actual desired value of the model
y'_i	Predicted value of the model
m	Number of input neurons for ELM; any real number greater than 1, cluster for FCM algorithm
n	Number of output neurons for ELM
I_m	Objective function for FCM algorithm
x_m	Input values for ELM
ω_m	Values of weight from input layer to hidden layer for ELM
β_m	Values of weight from hidden layer to output for ELM
O_j	Output value from ELM algorithm
u_{ij}	Degree of membership of x_i in the cluster j
A	Premise for FRBS
B	Consequence for premise A for FRBS
w	Vector of weight for MLP
x	Vector of inputs for MLP
y	Single output for MLP
b	Bias for MLP
φ	Non-linear activation function for MLP
ξ	Euclidean vector
t	Index of data in a given sequence
$x(t)$	New data item for SOM algorithm
$\alpha(t)$	Scalar factor for size correction of SOM
c	Index for SOM having smallest distance from $x(t)$ in Euclidean signal space
X_i, Y_i	Training data for SVM
R	Regularized risk function for SVM
C	Regularization constant for SVM, K-clusters for training Neural Network
L	Loss function for SVM
W	Regularizer for SVM
X_k	Randomly selected firefly for MFA
w_i	Adjusting coefficient for ANN-MFA
M_w	Weighting factor for ANN-MFA
M_b	Biasing factor for ANN-MFA
F_c	Set for K-prediction model
k	Number of clusters for K-prediction model
v_{ij}	Synaptic connection weight from the *i*-th input node x_i to the *j*-th neuron

n_h	Nodes in hidden layer for ANN-GA
$net_s^j(\cdot)$	j-th Static Activation Function
$net_d^j(\cdot)$	j-th Dynamic Activation Function
$net_o^l(\cdot)$	Activation function for l output neurons
m_d^j	Dynamic mean for j-th Dynamic Activation Function
σ_d^j	Dynamic standard deviation for j-th Dynamic Activation Function
ζ	Regularization parameter for GHSA-FTS-LSSVM
μ	Euclidean vector, RBF kernel function parameter
a_3	Approximate wavelet component for GNN-WT-GAF
d_1, d_2, d_3	Detailed wavelet components for GNN-WT-GAF

I. INTRODUCTION

Modern power system demands an uninterrupted supply of electricity to the load side. This requires a proper idea of predicting present and future load demand with the least amount of error. For achieving this goal, scientists and scholars have been trying to develop the most efficient and optimal state-of-the-art method for predicting the future demand for electricity consumption by a method known as load forecasting. Load forecasting is used to control several operations and decisions such as dispatch, unit commitment, fuel allocation, and off-line network analysis [1]. This gives the power utility company an idea about the future demand of the consumers and an ample amount of time to mitigate the difference between the generation capacity and load demand. Demand prediction minimizes the power generation cost and helps to establish an organized power system utility, especially because of the large expense pertaining to power generation. Different Machine Learning (ML) based techniques are widely used by many power and energy utility companies to predict the power or energy needed to equilibrate between generation and demand. In general, load forecasting can be termed as a technique for demand and supply management. However, it is a complex task requiring the analysis of various direct and indirect factors affecting the process. Even though there are several benefits of using load forecasting techniques, some challenges inhibit the accuracy of the methods. The process used for forecasting is convoluted and sometimes stochastic in nature. Weather-related variables further influence the data, complicating the forecasting. Therefore, the load at a given hour is not only dependent on the previous hour; rather it is affected by the data of the load consumption of the previous day, weather, demographic data, appliances number in the forecasting area, customer type, customer number, and econometric data, etc., [2]. Even in such varying circumstances, it is necessary to keep the load forecasting error as minimum as possible. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are the measures of prediction accuracy,

usually measured in percentage. The values for the evaluation of a certain algorithm are to be kept within a few percentage points for its viability in load forecasting.

Historical load data is the key component of a load forecasting model. Because to train the model, it has to learn the pattern of the electrical load data consumption. After that, the load data must be prepared for training purposes. The missing values and erratic data are corrected. Afterward, this electrical data is used to collaborate with data pertaining to other factors such as historical weather data, historical event data, etc. The accuracy of the forecasting is highly contingent upon such factors. All the aggregated data is then analyzed, and several models are selected for load forecasting. Among these models, the best with the most accuracy is chosen for implementation. There are several other factors that determine the accuracy of the load forecasting model. Most of them vary largely based on locations and equipment. These are also necessary to be brought under consideration while designing a precise load forecasting model. Such factors are usually considered as the input variable for the development of a model. Due to the unavailability of the desired data, it is sometimes difficult to consider all the factors. To develop a forecasting model in a certain region, the data of the factors are collected from respective sources, such as the weather data, which is collected from the local weather office, or the time factor data, which is collected from the calendar. The collection of data also faces several impediments due to the unavailability of proper data, anomalies in cases, missing values (more than 5%) in data, loss of data points are a few of the reasons to create problems in the data accumulation process. These problems have solutions that entirely depends on the case where the model is applied. The presence of one of these factors can cost the precision of the value. Hence, the possible solutions for addressing such challenges are being brought forth by researchers. Load forecasting with greater accuracy can be a great money-saving potential for electric utility corporations, whereas unwanted errors can lead to a great amount of financial and infrastructural loss. According to Haida and Muto et al. [3], both positive and negative forecasting errors can result in an increased generation cost of electricity. Only by 1% decrease in mean absolute percentage error (MAPE) and precise load forecasting have a consequential impact of 3-5% on the generation side by reducing the cost of generation about 0.1% to 0.3% [4]. In recent years, government and electric utility companies are showing great interest in renewable energy-based power generation. The progressing increase in renewable energy influences some great challenges. The stochastic nature of solar and wind power makes it challenging for the utility companies to match the load with the variable production of infrequent generation side. Researchers are trying to develop a new dispatch technique to replace traditional dispatch techniques that can operate with the addition of high-level intermittent energy sources like solar, wind energy, etc., [5]–[8].

This paper focuses on different single and hybrid methods based on ML. Two supervised learning methods,
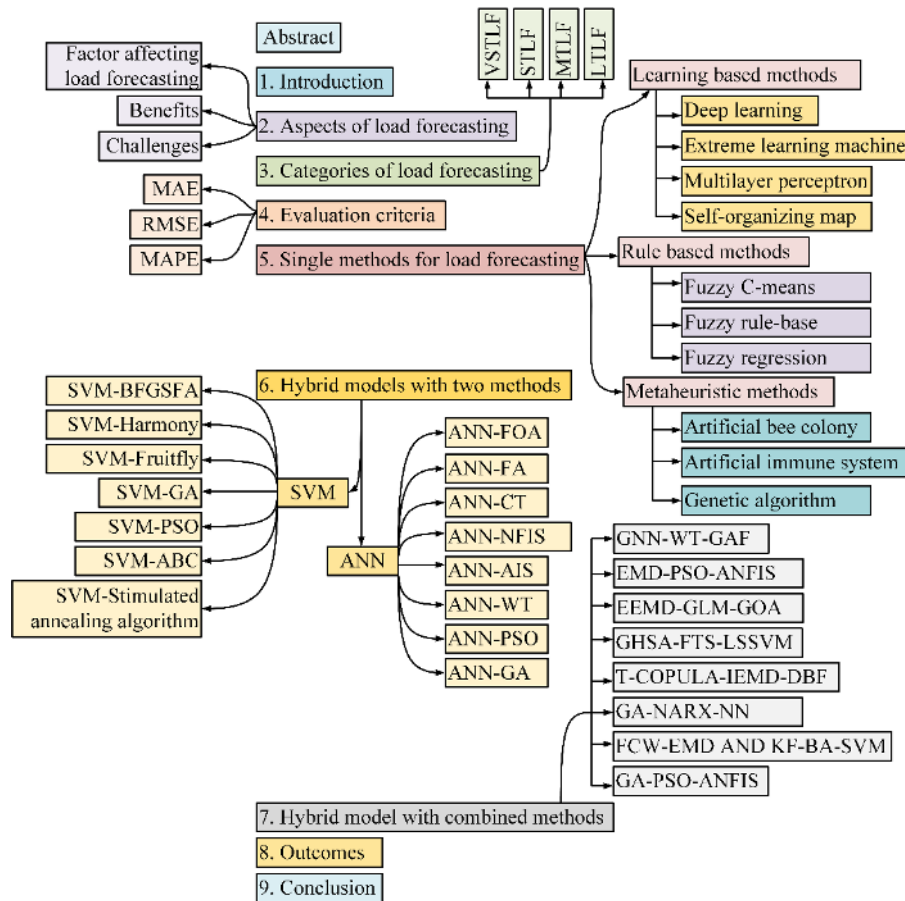
FIGURE 1. Information flow presented in the paper, with the explanation of single and hybrid models for load forecasting.

Support Vector Machine (SVM) and Artificial Neural Network (ANN) are widely used to design hybrid predictive models. In order to explore these two methods, different popular methods are delineated in the paper before discussing SVM and ANN-based hybrid algorithms. Some of these single methods are also optimized with either of these two algorithms. On the success of two method-based models, three or more single predictive models are being experimented upon to contribute to the improvement of forecasting accuracy. This paper reviews the current and previous work done on both the single methods and hybrid models for short-term load forecasting (STLF) utilizing the most used methods and prepares a comparative result of different algorithms used in distinct literature in order to present a comprehensive study of the load forecasting technology. Beginning with different aspects of load forecasting, the technique is classified based on the time horizon to scrutinize the algorithms' effectiveness in accordance with the prediction time. The efficacy of the utilized models is evaluated based on statistical criteria, to realize the significance of the models based on contingent factors. The hybrid models comprising of two or more methods are discussed to stipulate how the integration of algorithms can affect the forecasting technique. The flow of the information, as such, is shown in Fig. 1.

The rest of the paper is categorized as follows: Section II describes the factors, benefits, and challenges of load forecasting methodologies, section III categorizes the load forecasting methods based on the time horizon, section IV delineates the criteria for the evaluation of the accuracy of the models, section V reviews the basic principle and work done on single load forecasting model with the comparison of advantages and disadvantages; Section VI reviews the latest work done on hybrid models composed of two methods, where Part A discusses on the hybrid models optimized with SVM and Part B discusses on the hybrid models optimized with ANN. Section VII performs a comparative study on some hybrid models combining more than two methods. Section VIII enlists the significant findings of the paper. Finally, the conclusion is drawn in section IX with the indications of future research work.

II. ASPECTS OF LOAD FORECASTING
A. FACTORS AFFECTING LOAD FORECASTING
The empirical process of forecasting relies on several agents, which affects the precision of the process. Researchers need to choose the dependent factors carefully to acquire a correct prediction from the system. Time, weather and economic factors can be considered as the major effects to consider,
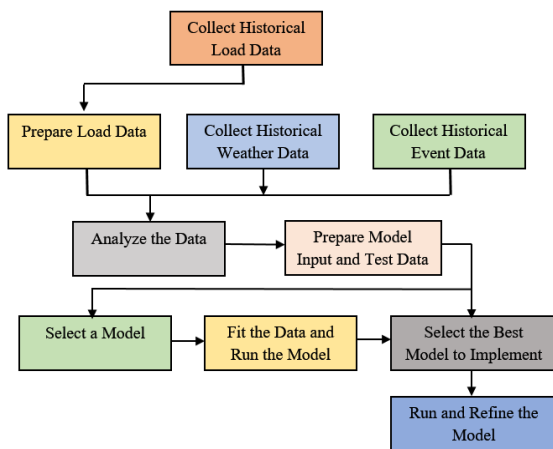
FIGURE 2. Flowchart of developing a load forecasting model; the load, weather and event data are collected to analyze and prepare the model to choose the best one for implementation.

although other major and minor factors affect every step of forecasting. Fig. 2 depicts the flowchart of the development of a load forecasting model, where the historical load, weather and event data are collected first from smart meters, sensors, data servers or other sources though various technologies and algorithms so that they can be prepared for analysis [9]. The selection of the model also requires significant attention as different algorithms use varying parameters as per these factors. Notable factors affecting data accumulation and model selection are briefed as follows:

- **Time Factor:** Time is the most important factor in load forecasting. From observing load curve of several different grid stations, Ruzic et al. [10] found that load curve has "time of the day" property along with "day of week", "week of month" and "month of season" property. This also indicates that not only the current data, but also the data of the previous days for a certain location also contribute to the accuracy of the prediction, which cannot be acquired by the reliability on the former data. Moreover, the timeframe to detect the load is also crucial as it defines the amount of data required for the process to run.

- **Weather Factor:** Weather is most independent variable in load forecasting domain, having its greater impact on domestic and agricultural consumers. The weather has a dominant effect on the behavior of the consumers. For example, in hot summer and cold winter, the electricity consumption goes up as the heating and cooling devices are turned on. This causes an increased demand of electricity at the warmest or coolest weather compared to the load demand during the days with average temperature. Also, sudden drop of temperature can cause less electricity consumption and thus causing overestimated load forecast. For the purpose, different models utilize the results from the weather forecast to predict the future load demand. The weather factors include temperature,

humidity, dew point temperature etc. Also, temperature humidity index (THI) and wind chill index (WCI) are broadly used by the utility companies. THI and WCI are the measurement of summer heat discomfort and cold stress in winter respectively.

- **Economic Factor:** Economic factors such as electricity price, management of load and degree of industrialization have important impact on system average load and maximum demand [11]. Moreover, customer behavior, change in tariff, description of appliances, population of the forecasting area, the age of the equipment, and employment levels play influential roles in determining the perfection of forecasting. For proper prediction in a particular area, usually for long term load forecasting, these factors must be brought under calculation as economic prediction considers public behavior to extrapolate load generation and demand, affecting data acquisition and model selection procedure [12].

B. BENEFITS
Load forecasting has its own benefits which attracted the researchers into this domain. Since the early age of the electricity generation, it was a burning question among the utility companies to determine the demand for the next hour, next day and even years, to make a balance between the limited resources and ever-increasing demand of electricity. Even though renewable energy sources have mitigated the resource management issue, the energy harvesting processes are still expensive and cumbersome. The electricity dispatching can also be optimized if the load demand can be estimated beforehand. That is when the term load forecasting came in. Apart from the economic and environmental perspective, there are other benefits of load forecasting. They are:

1. Understanding the future load demand helps the utility companies to plan, make economical viable decisions and minimize the risk for the companies. Also, decisions for future generation and transmission investments are made from load forecasting.

2. Load forecasting helps in planning the required resources for the future such as fuels required to operate the generation side as well as other resources that are important to ensure uninterrupted power to the consumers. This confirm economical generation and uninterrupted distribution of electricity.

3. To build a future generation plant, load forecasting helps in planning the plant size, location, capacity and type of the future generation plant. This gives a clear idea of the cost of transmission and distribution infrastructure as well as other associated loss.

4. By understanding the demand, load forecasting helps in deciding and planning for the maintenance of the power systems.

5. Finally, the load forecasting ensures maximum utilization of power plants by eradicating under generation

and over generation which in turns helps to reduce the usage of fossil fuel and reduce the carbon emission.

C. CHALLENGES

For many years, researchers are trying to develop accurate load forecasting model to improve the efficiency and revenues for the electrical generating and distribution companies. This has, so far, lead to the invention of many state-of-the-art methods. In recent years, researchers to investigate to improve the accuracy of these models, but several challenges are impeding the objective of the investigations. These challenges are the main hindrance in getting the best accurate model. Some of these difficulties to achieve the best forecasting model are mentioned below:

1. Electrical load forecasting is very much depended on weather. Unfortunately, sometimes the weather is not predictable and the forecasting may have a great error for sudden change in weather condition. Also, different regions in a big electrical system may have different weather conditions which affects the electricity demand. This creates a negative impact on the revenues.

2. Consumers in different region use different types of meters e.g. smart and traditional meters with different tariffs. Also, the usage behavior varies between these customers. The utility should have a clear understanding of the system utilization and develop separate forecasting model for each of the metering systems. Then add them up for the final forecast value. Otherwise there will be a huge forecasting error.

3. It is also difficult to get accurate data on consumption behavior due to sudden change in factors such as pricing and the corresponding demand based on such a price change.

4. Accurate load forecasting is difficult due to the complexity in fitting numerous complex factors which affect the demand for electricity into the forecasting models.

5. Sudden disturbances in the power system affects the regular load models. Unexpected faults and transients create exceptional data logs, which are to be considered separately for the design of the models, lest they should be distributed with the data equally, resulting in the design of a poor and unreliable forecasting system.

6. In addition, it is not easy to obtain an accurate demand forecast based on parameters such as change in temperature, humidity and other factors which influence the consumption of electricity.

7. Finally, the distribution company may suffer losses if they do not understand and decide on an acceptable margin of error in short term load forecasting.

The variation in the challenges motivated the designers to focus on different aspects of the model, collecting incidental data, parameterizing the algorithms and choosing the best model to make load forecasting as reliable as possible.

III. CATEGORIES OF LOAD FORECASTING

In terms of the time horizon, load forecasting can be categorized into four classes, on the basis of which different ML algorithms can be implemented: very short-term load forecasting (VSTLF), short-term load forecasting (STLF), medium-term load forecasting (MTLF) and long-term load forecasting (LTLF). VSTLF is popular for load forecasting from few seconds to few minutes. In VSTLF the load in the near future can be forecasted by the load in the past [13]. That is why temperature, economics and land use information can be optional. Extrapolating the recently observed load pattern to the nearest future is being used instead of modeling relationships between weather conditions, load, time and other affecting factors to the load. Methods for VSTLF are few, mostly including Autoregressive Moving Average Models, Artificial Neural Network, and Genetic Algorithm etc. STLF is used for a lead time from few minutes to few hours. It plays a vital role in system operations and it is the main source of information for all daily and weekly operations concerning generation commitment and scheduling [14]. As the load for long time horizons can be approximated from STLF, researchers are mostly interested in designing predictive models for the domain. For better short-term prediction modeling, it is required to have a proper knowledge on the factors affecting the load. Some of these factors are the weather conditions, the season, the type and time of the day of a specific area and many others. Integration between the factors and the load demand is the primary object to look for as the demand at any time of the day is different. MTLF is generally used for forecasting load from a few days to few months [15]. It is popular for forecasting load in seasonal changes such as winter or peak-summer etc. LTLF is used for lead time from few weeks to several years [16]. It takes into account the historical load and weather data, customer's number in categories, the characteristics of the appliances of the area etc. The economic factors are specially considered for long term forecasting methods. Table 1 shows the comparison between load forecasting methods based on time period, factors and application.

Among all the different methods, STLF is the most popular one. It plays a key role in the formulation of economic and secure operating strategies for the power system because of its inherent connectivity to other type of forecasts. STLF can be transformed into MTLF and LTLF by adding econometric variables to the STLF and extrapolating the model to the longer horizon. On the other hand, VSTLF model can be achieved from STLF by adding the loads of some preceding hours as a part of the inputs to the STLF model. Autocorrelation of the current hour load and the previous hour loads can be captured by short-term load forecasting. Also, the residuals of historical load can be collected and form a new series having the STLF as a base. Avery short term forecast can be obtained by forecasting the future residuals and adding them back to the short-term forecast. Process flow of conversion between STLF and LTLF, MTLF and VSTLF is shown in Fig. 3.

TABLE 1. Classification of load forecasting methods according to time period, factors and their application.

METHODS	TIME PERIOD	FACTORS			APPLICATION	
		ECONOMICS	TEMPERATURE	LAND USE		
VSTLF	Few Minutes	Not Mandatory	Not Mandatory	Not Mandatory	•	Distributions schedule
					•	Generation forecasting
STLF	Few Hours	Not Mandatory	Mandatory	Not Mandatory	•	Spinning reserve allocation
					•	Maintenance scheduling
MTLF	Few Days to Months	Mandatory	Simulated	Not Mandatory	•	Seasonal load forecasting
LTLF	>1 Year	Simulated	Simulated	Mandatory	•	Generation growth planning
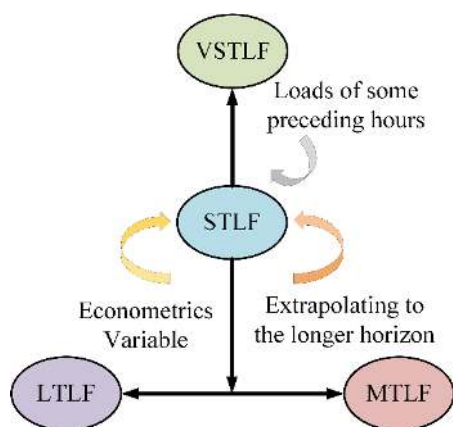
FIGURE 3. Process flow of conversion between STLF and LTLF, MTLF, VSTLF with the transaction of economic variables, statistical process and time.

FIGURE 4. Flowchart of simple short-term load forecasting method, showing the modeling and extrapolating process driven by weather, load data and weather forecast values for load prediction.

As depicted in Fig. 4, the simple process flowchart of STLF takes weather and load history as the input of the modeling process to model the extrapolating process with the

FIGURE 5. Most used single methods for short-term load forecasting method, some of which are used for other medium and long term predictions as well.

accumulation of the weather forecasting data. The forecasting data is then considered for the minute or hourly prediction of the load. There are many STLF techniques that are designed for the model. Some of them are time series analysis, regression analysis, artificial neural networks (ANN), support vector machine (SVM), fuzzy logic (FL), genetic algorithms (GAs) and hybrid methods etc., as listed in Fig. 5. Due to their self-adaptive mechanism and the features of mimicking the intelligent behaviors in complex and continually changing behavior.

Computation Intelligence (CI) methods are widely used in the current research works. CI method is denoted as the potentiality of computer algorithms to intuitively learn a particular task from the trial and error perceptions form the available data from the past and predict accurately in the future based on the learning. The unique feature of CI methods is related to their ability for autonomous operation without requiring any complex mathematical formulations or

TABLE 2. Evaluation criteria formula.

Criteria Name	Formula
Mean Absolute Error (MAE)	$\dfrac{\sum_{i=1}^{n}\lvert y_i' - y_i\rvert}{n}$
Root Mean Squared Error (RMSE)	$\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i' - y_i)^2}$
Mean Average Percentage Error (MAPE)	$\dfrac{1}{n}\sum_{1}^{n}\dfrac{(y_i - y_i')}{y_i}$

quantitative correlations between the inputs and the output. The hybridized version of CI models has achieved significantly better performance than their single model counterparts [17].

IV. EVALUATION CRITERIA
Several criteria are used as the evaluation indices of load forecasting techniques to justify how correct the methods are to predict the actual load values. Different researchers have depended on different statistical metrics to quantify the accuracy of their model and with the passage of time new statistical metrics are coming into play such as metrics for probabilistic load forecasting. The probabilistic load forecasting literature is still developing as it shows promising academic value and wide adaptation in the industry. Table 2 shows the most popular static metrics among the researchers around the world.

Here, n is the number of samples, y_i' is the predicted value of the model and y_i is the actual desired value. Each of these metrics has its own advantages and disadvantages. RMSE provides a loss function of second degree but emphasizes greater errors than the small ones. MAE is unambiguous and it can measure average error naturally. MAPE does not depend on scale and it can be applied easily to both high and low volume products. But due to differential penalty often it can lead to biased forecasting. The weaknesses of MAPE, such as difficulties in handling small and zero denominators, are not very relevant for traditional load forecasting problems, because the load at the aggregated level is rarely zero or approaching a very small number [18].

Also the value of these metrics are different for different datasets and parameters. Therefore it is quite difficult to compare the results of different techniques. Also, there is no such task where all the methods are experimented in a single dataset, to find the comparison between them. In this research work, the best accuracy of each discussed method are tabulated in the subsequent sections from different forecasting methods.

V. SINGLE METHOD FOR LOAD FORECASTING
Using parametric, nonparametric and artificial intelligence based single methods, load forecasting is performed for very short time. ANN, Local Fuzzy Reconstruction Method, and specific regression and statistical models are being used to

forecast for a few seconds to a dozen of minutes [19]. For STLF, the versatility of algorithms such as Expert Systems, Time Series Analysis, Similar Day Look Up Approach, SVM and FL augments as shown in Fig. 5. Single methods used for STLF are somehow reliable for MLTF as well. Grey model, Adaptive Neuro Fuzzy Inference System (ANFIS), and Wavelet Transform (WT) and several distinguished methods are added to the list for LTSM along with those used for MLTF [20]. For the employment of such methods, a clear concept of the exercised terminologies are necessary, which are provided and compared with respect to the evaluation criteria in the subsection that follows.

A. LEARNING BASED METHODS
1) DEEP LEARNING (DL)
The term Deep Learning (DL) was first introduced by Dechter [21] in 1986. Apparently, after a gradual improvement in the algorithm and context, it is now renowned as Deep Neural Network (DNN). A neural network is a shallow structure consisting of an input layer, a hidden layer and an output layer. However, a deep learning system architecture consists of more layers than a traditional three-layered multilayer perceptron (MLP). The networking structure of deep learning is the best possible simulation of human cerebral cortex, where it mimics the human brain function [22]. Due to its feature of modeling nonlinearity, DL is widely used in various forecasting applications [23]. It can represent the complex high-dimensional functions and hyper variable functions, which deteriorates the computational complexity of the forecasting models. One major drawback of this system is the over-fitting issue due to the requirement of large number of layers for precise output. Due to its algorithmic complexities, DL often requires a huge amount of runtime. There are several DL methods including convolution neural network (CNN), deep belief networks (DBN) and deep auto encoder (DAE) etc. Ryu *et al.* [23] proposed two different DNN models to learn complicated relations between weather variables, date and previous consumptions for individual customers.

This DNN model is later used to produce a day-ahead forecast of 24 hours load profile from the past data observations. He *et al.* [24] proposed a model to forecast the hourly load of a power grid. Their model combines the co-movement analysis from Copula model with layer-wise pre-training-based deep belief network. A comparative analysis between Support Vector Regression (SVR), Neural Network (NN), Extreme Learning Machine (ELM) and classical DBN in both day-ahead and week-ahead forecasting shows better result in the proposed semi-parametric data-driven method. The major over-fitting issue of DL is addressed in a study of Shi *et al.* [25]. By increasing the diversity and volume of the data, they were able to address the over-fitting issues of DL. Their proposed novel pooling-based deep recurrent neural network (PDRNN) batches a group of customer load profiles into a pool of inputs. The proposed method consists of two stages: 1) load profiles pooling, and 2) household STLF with Deep-Recurrent Neural

Network (D-RNN). This method has a prerequisite of sufficient diversity in customer load for learning common sharing uncertainties such as weather conditions and pertaining factors. Testing on 920 smart metered customers from Ireland, the proposed method outperforms ARIMA by 19.5%, SVR by 13.1% and classical D-RNN by 6.5% in terms of RMSE. A framework based on resident behavior [26] was designed with a long-short term memory (LSTM, a variant of RNN) based DL, which took appliance consumption sequence into consideration. A constant nature of the power grid network system is its high variability and volatility. To mitigate this challenge and reduce the errors of the forecasting of the electric parameters, DBN is combined with Copula Model [27] and bidirectional Recurrent Neural Network [28], [29] for day and week ahead forecasting. An LSTM based hybrid DL method was tested by Motepe *et al.* [30] for South African distribution network, showing an improved performance, considering the inclusion of temperature data. Deng *et al.* [31] devised a deep multi-scale CNN (MSCNN) with time cognition and a self-designed time coding algorithm, which outperformed recursive multi-step LSTM, direct multi-step MSCNN and the direct multi-step gated CNN by MAPE of 34.73%, 14.22% and 19.05% respectively. An improved DBN especially for Demand Side Management was designed by Kong [32] for STLF, outperforming autoregressive integrated moving average (ARIMA), Least Square SVM and conventional DBM with MAPE and RMSE of 3.864 and 341.601 respectively. RNN and its different approaches are widely used for short-term residential load forecasting [33]. LSTM was further integrated with Gated Recurrent Unit (GRU) for hybrid distribution feeder LTLF by Dong and Grumbach [34]. The model was scrutinized for an urban grid in West Canada, which exceeded the performance of ARIMA, bottom-up and feed-forward NN.

### 2) EXTREME LEARNING MACHINE (ELM)

Huang *et al.* [35] reviewed that learning speed of feed-forward neural network is far slower than required and the main reasons behind this are: (1) the slow gradient-based learning algorithms that are extensively used to train neural networks, and (2) all the parameters of the networks are tuned iteratively. To overcome this problem, the authors proposed a new learning algorithm called extreme learning machine (ELM) for single-hidden layer feed-forward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. ELMs are feed-forward neural networks with a single layer of multiple layers of hidden nodes. These hidden nodes can be randomly assigned and never updated (i.e. they are random projection but with nonlinear transforms), or can be inherited from their ancestors without being changed. Another interesting feature of this method is that these hidden layers need not to be tuned.

Fig. 6 shows the structure of ELM, depicting m number of inputs denoted as $x_1, x_2, \ldots x_m$. These inputs traverse through n hidden layers. Each values from the input neu-
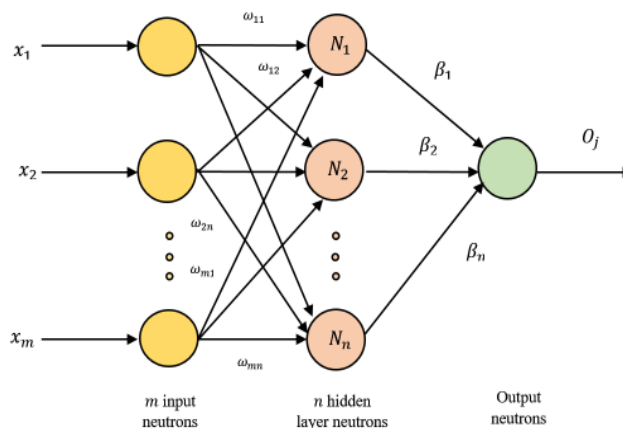
FIGURE 6. Structure of extreme learning machine, with m input neurons choosing between n hidden layer neurons proceeding towards the forecasting of output O_j [36].

rons are assigned a different weights $\omega_1, \omega_2, \ldots \omega_n$ from input to the hidden layer and linear weights $\beta_1, \beta_2, \ldots \beta_n$ from the hidden layer to acquire the predicted output O_j. Its learning speed can be thousands of times faster than traditional feed-forward neural network learning algorithms like back-propagation (BP), which can obtain better generalization performance [35]. Also the ELM tends to reach the solutions straightforward without facing issues like local minima, improper training rate and over fitting etc. Li *et al.* [37] proposed a novel ensemble method for short-term load forecasting where wavelet transform, ELM and partial least squares regression are integrated. From different combinations of mother wavelet and number of decomposition levels the individual forecasters are derived. A parallel model consisting of 24 ELMs is invoked to predict the hourly load of the next day, for each sub-component from the wavelet decomposition. The individual forecasts are then combined to form the ensemble forecast using the partial least squares regression method. This proposed method has been tested using data from two electric utilities for 1 hour and 1 day ahead load forecasting and the output result have showed better forecasting accuracy than other state-of-the-art models. Ertugrul [38] proposed a new model incorporating ELM with RNN, naming the model Recurrent Extreme Learning Machine (RELM). RNN shows better results in the forecasting dynamic systems compared to the feed forward ANN model. This model shows great result as the training time is much lower and this method can be used for real-time dynamical systems for load forecasting. Zhang *et al.* [39] worked with the short-term load forecasting emerged on ELM under the supervision on Improved Gravitational Search Algorithm (IGSA), which is a combination of Particle Swarm Optimization and Gravitational Search Algorithm, where it is used to search the optimal set of input weights and hidden biases for the ELM. Li *et al.* [40] proposed a STLF method based on the improved extreme learning machine which is the capable of autonomously selecting the number of hidden neurons according to the group number of input samples, which makes
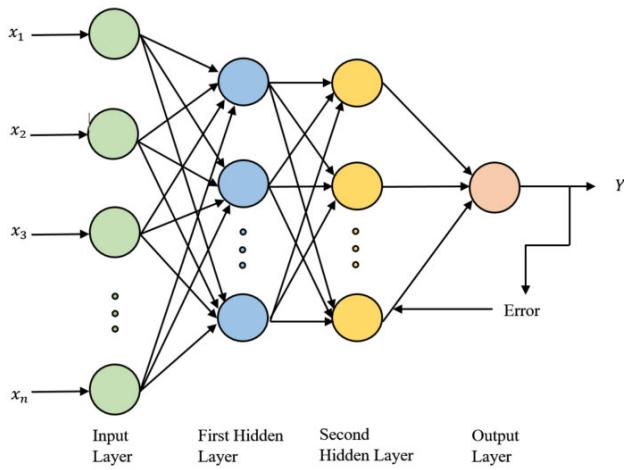
FIGURE 7. Structure of an ideal MLP model with two hidden layers, taking the output error into consideration for enhancing the accuracy.

the training error close to zero, and the test error is as tiny as possible. Chen *et al.* [41] proposes a novel short-term load forecasting method which is based on empirical mode decomposition (EMD) and ELM. EMD is an empirical approach to obtain instantaneous frequency data from non-stationary and non-linear data sets. The empirical mode decomposition is utilized to decompose the load series for capturing the complicated features of the electric load and to de-noise the data [42]. This method is tested on the state of New South Wales, Victoria and Queensland in Australia for half hourly electric load forecasting and the results clearly showed great improvement.

3) MULTILAYER PERCEPTRON (MLP)

A perceptron is an algorithm that classifies input by separating two categories with a straight line. In other words, it is a linear classifier. Based on several real-valued inputs by forming a linear combination using its input weights, a perceptron generates a single output, **y** as shown in the following equation.

$$ y = \varphi \left(\sum_{i=1}^{n} w_i x_i + b \right) = \varphi(w^T x + b) \quad (1) $$

where, **w** denotes the vector of weights, **x** is the vector of inputs, **b** is the bias and φ is the non-linear activation function.

An MLP is basically a deep, artificial neural network (ANN). Fig. 7 demonstrates the structure of an ideal MLP, having a three-layered structure. One is input layer which receives the signal, an output layer that makes a prediction about the input and in between those two, an arbitrary number of hidden layers that are the core computational machine of the MLP. It has the ability of manipulating the input space by adjusting weight matrices continuously between layers until the error between target value and the predicted value is minimized.

Ferreira and Da Silva [43] developed two nonparametric procedures for solving the problems of NN structure

and input selection for STLF. Their proposed model can improve the outcome of MLP and RBFs. Ding *et al.* [44] compared between the Naïve model and NN model using MLP for STLF. The investigation shows that accuracy of the MLP model is 4.7% better than the Naïve model. Bokingkito *et al.* [45] used Multilayer Perceptron Neural Network model for week ahead load forecasting. The sigmoid activation function in resilient propagation showed the most efficient and least network error in the training. Interestingly, in the study of Kuo and Huang [46], the proposed deep neural network named DeepEnergy which outperformed MLP, RBF and several other popular machine learning technology. Due to the effectiveness of MLP in MTLF, Askari and Keynia [47] combined two search algorithms, i.e. Particle Swarm Optimization (PSO) and improved Ant-Lion Optimiser to design an MLP based model to solve the MTLF problem.

4) SELF-ORGANIZING MAP (SOM)

The Self-Organizing Map also known as Kohonen network was initially developed by Kohonen [48] in 1982. It is a computational method for the analysis and visualization of high-dimensional data, a type of ANN that is trained by unsupervised learning to generate low-dimensional, discretized representation of the input space of training samples, called a map. SOM apply competitive learning as opposed to error correction learning e.g. BP with gradient descent and use a neighborhood function to preserve the topological properties of the input space. Consider first data items that are n-dimensional Euclidean vectors

$$ x(t) = [\xi_1(t), \xi_2(t), \ldots, \xi_n(t)] \quad (2) $$

where, t is the index of the data item in a given sequence.

Let the i-th model be represented by the following equation,

$$ m_i(t) = [\mu_{i1}(t), \mu_{i2}(t), \ldots, \mu_{in}(t)] \quad (3) $$

where, t denotes the index in the sequence in which the models are generated. This sequence is defined as a smoothing-type process in which the new value $m_i(t+1)$ is computed iteratively from the old value $m_i(t)$ and the new data item $x(t)$ as:

$$ m_i(t+1) = m_i(t) + \alpha(t)hc_i(t)[x(t) - m_i(t)] \quad (4) $$

where, $\alpha(t)$ is a scalar factor that defines the size of the correction; its value decreases with the step index t. The index i refers to the model under processing, and c is the index of the model that has the smallest distance from $x(t)$ in the Euclidean signal space. The factor $hc_i(t)$ is a kind of smoothing kernel, also called the neighborhood function. It is equal to 1 when $i = c$ and its value decreases when the distance between the models mi and mc on the grid increases. Moreover, the spatial width of the kernel on the grid decreases with the step index t. These functions of the step index, which determine the convergence, must be chosen very exquisitely.

Fan *et al.* [49] used traditional SOM to learn time-series load data with weather information as parameters and to

improve the accuracy of the prediction. The dataset used in this study consist of the daily peak electrical consumption and weather data in New York City and Long Island from July 1, 2001 to September 31, 2004. An extension of SOM algorithm based on error-correction rule is also used. The peak load is generated by averaging the output of all the neurons. Their model achieved minimum number of MAPE value 1.93% using 15 × 15 number of neurons. In their study, López *et al.* [50] showed the effect of SOM neural networks in load forecasting. They also performed a deep and through analysis of real world prediction. In this study, data from Spain energy consumption from 2001 to 2010 has been used to validate the model. The proposed model forecasts daily market load with 2.32% MAPE value. Li *et al.* [51] proposed a localized Bayesian-Regularization NARX (Non-linear Autoregressive models with Exogenous Input) Neural Network model combining with Self-Organizing Mapping, where SOM is utilized to extract the meteorological distribution and K-means is used to cluster the data according to the nearest mean. They also used the Bayesian-Regularization BP algorithm to train the target network structure which eventually improved the predictive ability of the localized NARX neural network. The proposed model is tested in Sydney using half hourly power system load and electricity prices data. The results showed a great accuracy in prediction which can prove beneficial both economically and socially.

B. RULE BASED METHODS

1) FUZZY C-MEANS (FCM)

The Fuzzy C-means (FCM) clustering algorithm was initially developed by Dunn [52] and was later developed by Bezdek *et al.* [53]. FCM is a method of clustering which allows one piece of data to belong to two or more clusters. Clustering is a popular and widely used mathematical tools that attempts to discover structures or certain patterns in a dataset and the objects inside each cluster showing a certain degree of similarity. Among the clustering algorithms, one of the most popular algorithms, Fuzzy clustering algorithms and Fuzzy Set theory was first proposed by Zadeh [54]. He gave an idea of the uncertainty of belonging which was described by a membership function and the central idea in fuzzy clustering which is the non-unique partitioning of the data into a collection of clusters. FCM algorithm is based on minimization of the following objective function:

$$I_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \tag{5}$$

where, m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster m, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster.

Radial Basis Function (RBF) network is superior to BP network because the ability to approach nonlinear function and the convergent speed are better the ones of BP network model [55]. Zhu and He [56] overcame the slow convergence speed and local minima problem of BP network algorithm
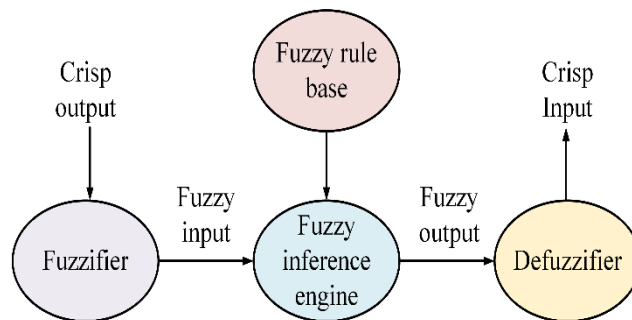
FIGURE 8. Algorithmic structure of fuzzy rule base system, showing the conversion of crisp to fuzzy values through fuzzification and defuzzification process.

by introducing the application of FCM based RBF model to short term load forecasting problem. The proposed algorithm was tested on the actual power load data and the results showed a good average percentage error of 4.04%. Yi *et al.* [57] proposed an ultra-short time forecasting based on FCM algorithm to increase flexible process capability of power systems. FCM algorithm was used to calculate the daily load change rate. This method showed more accurate and predictable result in the practical application of power grid company where the error analysis index showed much better result than the traditional prediction algorithm. Clustering is equally used for LTLF [58] and residential load forecasting [59] due to its improved fluctuation management algorithm.

2) FUZZY RULE BASE SYSTEM (FRBS)

Fuzzy rule-base system (FRBS) is one of the great applications of fuzzy set or fuzzy logic. Fuzzy logic systems address the imprecision of the input and output variables by defining fuzzy numbers and fuzzy sets which that can be expressed in semantic variables. Fuzzy rule-based approach is based on verbally formulated rules which is overlapped throughout the parameter space. They use numerical interpolation to handle complex non-linear relationships. Fuzzy rules are linguistic IF-THEN- constructions that have the general form "IF *A* THEN *B*", where *A* and *B* are (collections of) propositions containing linguistic variables.

A is called the premise and *B* is the consequence of the rule. In effect, the use of linguistic variables and fuzzy IF-THEN-rules exploits the tolerance for imprecision and uncertainty. In this respect, fuzzy logic mimics the crucial ability of the human mind to summarize data and focus on decision-relevant information. As shown in Fig. 8, the raw output, called the crisp output is inserted in a process "Fuzzyfier" to 'Fuzzify' the crisp values for generating the fuzzy input. The Fuzzy rule base are determined by the user to the Fuzzy inference engine or software as per the requirement of the forecasting system, From the Fuzzy input, the inference engine follow the conditions to generate Fuzzy output, which are then converted to a crisp input by a "Defuzzifier" for real world application.

FRBS is the combination of two ideas; one is knowledge and another is reasoning. The approximate reasoning is the very popular fuzzy set that has multi-valued logic systems was introduced by Lukasiewicz in the 1930 and modified by Zadeh in the 1960 [54]. Ranaweera *et al.* [60] did a comprehensive exploratory investigation of the application of a Fuzzy Logic model for STLF problem. The proposed model showed a great forecast with a MAPE value below 2.3%. One of major problems faced when designing a fuzzy model is the finding of an optimized fuzzy rule base. Generally it is done by a process called trial-and-error which is time consuming and arduous. But Kang *et al.* [61] presented an approach to the evolutionary design of an optimum fuzzy rule base for modeling and control where evolutionary programming is used to simultaneously evolve the structure and the parameter for a fuzzy rule. The method's capability in simultaneous handling of quantitative or qualitative information and uncertainties have been discussed by Khosravi *et al.* [62]. They have proposed the application of IT2FLS for STLF. Later, they have developed a model using Takagi-Sugeno-Kang fuzzy inference systems (TSK-FIS) to develop IT2-TSK-FLS hybrid algorithm for load forecasting. The experiment conducted on actual load data shows precisely approximated future load demands with an acceptable accuracy. Hassan *et al.* [63] presented a method of interval type-2 fuzzy logic systems (IT2FLS) for load forecasting where extreme learning machine (ELM) is used to tune the parameter of IT2FLS for chaotic and nonlinear data. This method used multi-inputs for proper input-selection along data pre-processing. To select the influential inputs a partial autocorrelation analysis was utilized. The time-delays of the dataset which have cabalistic coefficients are selected as input to the model. Ali *et al.* [64] proposed fuzzy logic model for LTLF based on the weather parameters (temperature and humidity) and historical load data. This model comprises the steps for data collection, constructing fuzzy interface, Fuzzifying input and output, assigning membership function, setting up Fuzzy rule base, building fuzzy logic models and simulations, and analyzing errors. The fuzzy logic model forecast a year-ahead load with a MAPE of 6.9% and efficiency of 93.1%. Another TSK based method used Fuzzy model across the entire domain, improving MAPE by 0.13% compared to seasonal autoregressive integrated moving average (SARIMA) and Gustafson-Kessel clustering [65].

3) FUZZY REGRESSION (FR)
Regression-based methods such as autoregressive (AR) models, autoregressive moving average (ARMA) models, and autoregressive integrated moving average (ARIMA) models are usually applied to make short-term load forecasting [66]. Fuzzy regression is introduced in order to overcome some of the limitations of linear regression, such as the vague relationship between the dependent variable and the independent variables, insufficient numbers of observations, and hard-to-verify error distributions. The fundamental difference between the assumptions of the two techniques relate to

the deviations between the observed and estimated values: linear regression assumes that these values are supposed to be errors in measurement or observations, while fuzzy regression assumes that they are due to the indefiniteness of the system structure [18]. Many interesting models based on fuzzy regression for electric load forecasting are already described in the available literature. Song *et al.* [67] developed the forecasting model for short-term prediction using Tw-based fuzzy arithmetic operations. In [68], several fuzzy linear regression models were compared in forecasting performance on electric load data. For oil consumption forecasting the FLR served also very well [69].

More complex model for time series forecasting with several pre-processing and post-processing procedures, including fuzzy regression, was proposed in [70]. This paper presents a fuzzy polynomial regression method with data selection based on Mahalanobis distance incorporating a dominant weather feature for holiday load forecasting. Selection of past weekday data relevant to a given holiday is critical for improvement of the accuracy of holiday load forecasting. In the paper, a data selection process incorporating a dominant weather feature is also proposed in order to improve the accuracy of the fuzzy polynomial regression method. The dominant weather feature for selection of historical data is identified by evaluating mutual information between various weather features and loads from season to season [71].

This application-oriented paper proposes a fuzzy interaction regression approach to STLF. Through comparisons to three models (two fuzzy regression models and one multiple linear regression model) without interaction effects, the proposed approach shows superior performance over its counterparts [72].

A possibilistic linear function can be defined as:

$$Y = A_1 X_1 + A_2 X_2 + \ldots . A_n X_n = Ax \tag{6}$$

where, x_n is non-fuzzy. A_n is a symmetric fuzzy number denoted by $(\alpha_i, c_i)_L$, with α_i as the center and c_i as the spread. In this paper, we assume that the reference function $L(x) = \max(0, 1 - |x|)$. The type of fuzzy parameter Ai is a symmetrical triangular fuzzy number

$$\mu_{Ai}(ai) = L((a_i - \alpha_i)/c_i) \tag{7}$$

where $c_i > 0$. The possibilistic linear function $Y = Ax$ is obtained by the following membership function:

$$\mu_Y(y) = \begin{cases} \mathrm{L}((y - x^T a_i / c^T |x|), & x \neq 0 \\ 1, & x \neq 0, y = 0 \\ 0, & x = 0, y \neq 0 \end{cases} \tag{8}$$

where, $|x| = (|x_1|, |x_2|, \ldots, |x_n|)^T$.

Identification of the parameters of the fuzzy linear regression model can be formulated as a linear programming problem:

$$\min_{\alpha, c} J(c) = c^T |x|$$

$$s.t \; y_i \leq \left| L^{-1}(h) \right| c^T |x_i| - x_i^T \alpha,$$

$$- y_i \le \left| L^{-1}(h) \right| c^T |x_i| - x_i^T \alpha,$$
$$c \ge 0,$$
$$i = 1, 2, \ldots \ldots, N, \tag{9}$$

where, h is the threshold to control the width of the spread, and $0 \le h < 1$.

Vantuch and Prílepok [73] proposed an innovative algorithm entitled as ensemble of fuzzy linear regression (EFLR) and it bases on fuzzy linear regression combined with boosting mechanism. The fuzzy linear regression is optimized making use of multi objective optimization. The original data of electric load patterns are involved in order to develop and evaluate a load forecasting model as an experimental application of EFLR. The comparison of EFLR with basic fuzzy linear regression revealed improvement of more than 2% in all measures which proves the necessity of ensemble-based approach in the fuzzy linear regression. Luy *et al.* [66] studied on handling the problem that is caused by the growing knowledge base, and improves the load forecasting performance of fuzzy models through nature-inspired methods. The proposed models have been optimized by using ant colony optimization and genetic algorithm (GA) techniques. The training and testing processes of the proposed systems were performed on historical hourly load consumption and temperature data collected between 2011 and 2014. The results show that the proposed models can sufficiently improve the performance of hourly short-term load forecasting. The mean absolute percentage error (MAPE) of the monthly minimum in the forecasting model, in terms of the forecasting accuracy, is 3.9% (February 2014). The results show that the proposed methods make it possible to work with large-scale rule bases in a more flexible estimation environment. Although fuzzy regression has been tried for STLF for about a decade, most research work is still focused at the theoretical level, leaving little value for practical applications. A primary reason is that inadequate attention has been paid to the improvement of the underlying linear model [72].

C. METAHEURISTIC METHODS

1) ARTIFICIAL BEE COLONY (ABC)

Artificial Bee Colony is a swarm based meta-heuristic algorithm that was first developed by Karaboga [74] in 2005. Karaboga developed this model based on the model proposed by Tereshko and Loengarov [75] for the foraging behavior of honey bee colonies. The minimal model of forage selection that leads to the emergence of collective intelligence of honey bee swarms consists of three essential components: food sources, employed foragers, and unemployed foragers. It is an optimization tool which is developed by the intelligent behavior of honey bees where individuals called food positions are modified by the artificial bees with time and the aim of the bees is to discover the food sources with high nectar amount and finally the one with the highest nectar. Honey bee colony is the combination of employed bees, onlookers and scouts where number of employee bee same as the food
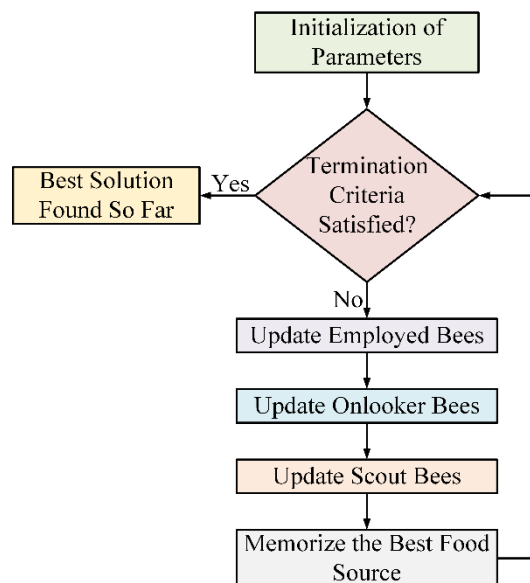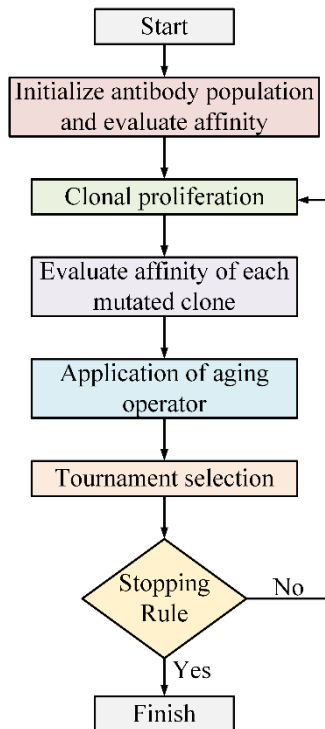

FIGURE 9. Algorithmic flowchart of artificial bee colony, the employed, onlooker and scout bees are updated to save the best food source till the termination criteria is satisfied.

source. The employee bee first calculates the nectar amount and gives this information to the onlooker's bee. Onlookers bee find the best solution of source, and then the scout bees go to collect the honey [74]. ABC algorithm was developed based on this basic working principal of honey bee where first half swarm is consider employee bee and least of the half is Onlookers bee. Employed bees perform the exploitation process in the search space and the scout's control the exploration process [76]. It has the great ability to get out of a local minimum problem. Fig. 9 demonstrates the flow of command for the ABC algorithm. After the initialization of the parameters, employed, onlooker and scout bees are updated to remember the best source of food. The search iterates until it meets the termination criteria to choose the best solution. One of the great limitations of this algorithm is convergence performance of ABC for local minimum to be slow [77]. Safamehr and Rahimi-Kian [78] developed cost efficient method to solve the proposed two-objective optimization problem of a micro grid using intelligent demand-response program such as ABC. Çevik *et al.* [79] proposed swarm intelligence viz. ABC and PSO algorithm for load forecasting on smart grids. They used previous data of four years (2009-2012, Turkey) for training and validation where three years data were used for creating model and one year data were used for validation. Load profile divided into four parts as winter, spring, summer and autumn.

#### 2) ARTIFICIAL IMMUNE SYSTEM (AIS)

The Artificial Immune System (AIS) was first introduced in the mid-1980s with articles authored by Farmer *et al.* [80] and later developed by Bersini and Varela [81] on immune networks. According to Rowe [82], the IS can function as "second brain", as it can generate responses to new and

FIGURE 10. Algorithmic flowchart of artificial immune system. After starting the antibody population and affinity evaluation, the clonal operation is conducted until the stopping criteria is met.

novel networks. Artificial Immune System is a mathematical and computer modeling of immune systems, or the abstraction of immunology-related principles into algorithms to address systematic goals. Immune System (IS) has an adaptive response that enables it to learn protein structures that characterize pathogens it encounters, and remember those structures [83]. This helps the IS to response to the same pathogens quickly and efficiently in future [84]. Development of AIS can be seen as having two target domains: the provision of solutions to engineering problems through the adoption of immune system inspired concepts; and the provision of models and simulations with which to study immune system theories [85]. However in the mid-1990s, AIS became a field in its own right and initial articles were published by Forrest *et al.* [86] and Kephart [87]. Also, in an article, Dasgupta [88] made an elaborate overview on Artificial Immune Systems and its application. According to Fig. 10, an initialization of antibody population led the algorithm to perform clonal proliferation. Affinity of each mutated clone is evaluated. The tournament for further proliferation is selected based on the aging operation. The iteration proceeds until the stopping rule is met. Yong *et al.* [89] propounded a load forecasting method using AIS which is the combination of Immune Network Regulation and Immune Programming.

AIS algorithm is used to sustain the diversity of population and defeated the deficiency of premature phenomenon so that it enhances the speed of searching and precision of optimization. This model showed the MAPE of 2.038%. A load note of a certain 10 KV distributing net in Beijing and its data
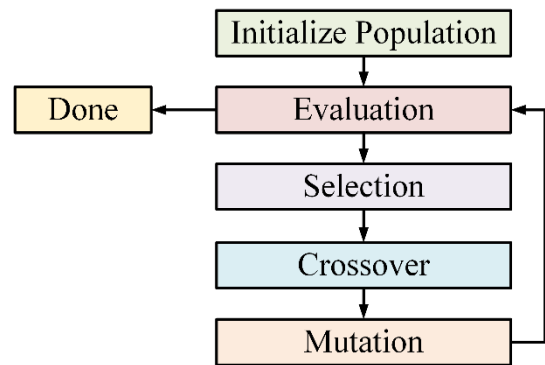
FIGURE 11. Algorithmic flowchart of genetic algorithm, showing the order of operation to generate a new population by selection, crossover and mutation to be fed back for evaluation.

are processed as the example. The proposed method of this study has been validated by the historical load data used in this study.

Dudek [90] proposed an electricity load forecasting method to mitigate gap between the demand with generation where AIS is used to train to recognize the antigens. It gives the one-day ahead power system load forecasting. The complete learning process of antibody illustrates overlapping clusters of same antigens where solving the problem is considered as antigen and solution of the problem can be considered as antibody.

3) GENETIC ALGORITM (GA)

Inspired by Charles Darwin's theory of natural evolution, Genetic Algorithms (GAs) are adaptive heuristic search algorithm which is used for solving both constrained and unconstrained optimization problems. GAs are frequently used to produce high-quality solutions to optimization and search problems based on bio-inspired operators such as mutation, crossover and selection. This method was initially discovered by Holland [91] of University of Michigan. Three main types of rules are used by Genetic Algorithm at each step to create the next generation from the current population such as:

- Selection rules: select the individuals known as parents which contribute to the population of the next generation.
- Crossover rules: combine two parents to form children for the next generation.
- Mutation rules: apply random changes to individual parents to form children.

Fig. 11 presents the order of the algorithm. The data or the population, upon initialization, proceeds to the evaluation process, where specific populations are selected for a crossover to generate a new population with combined characteristic of the previous population. After crossover and mutation, the new population is fed back for evaluation, until the evaluation meets the cherished result.

Ghareeb and El Saadany [92] presented a new variant of genetic programming, namely Multi-Gene Genetic Programming (MGGP) for STLF. In MGGP, each individual of the

TABLE 3. Comparative study among single load forecasting method.

Methods	Advantage	Weakness
DL	• Layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure. • Reduce the computational complexity. Provides better initial values of the network	• Low empirical coverage probability. • It is not scalable • Slower convergence and also the training process may get trapped at bad local optima
ELM	• Extremely fast training, • Good generalization, and Universal classification capability • Ease of implementation, minimal human intervention	• ELM suffered from both the uncertainty and generalization degradation problem
Fuzzy C Means	• Widely used in pattern recognition. • Error analysis index is much better than other load forecasting method. • Accurate and practicable.	• Need to consider temperature, humidity, weather type. • Long computational time • Sensitivity to the initial guess (speed, local minima)
Fuzzy Rule Base	• There are no design parameters in the estimation of the parameters which eliminates the need for trial and error and minimize the development time of the fuzzy model. • Universal approximation capability • Possess the capability of handling quantitative/qualitative information and uncertainties.	• Difficulties to identify of an optimal fuzzy rule base. • The computational complexity hinders their widespread use in practical applications.
MLP	• MLPs are very good in modeling and patter recognition. • Provide better accuracy in highly nonlinear problem. • MLP has the highest classification accuracy and insignificant errors.	• Convergence can be slow • Local minima can affect the training process
SOM	• Complex sets of pattern recognition capability. • Requires a very few numbers of training samples. • It allows a better interpretation of results	• Inputs need to be vectors • SOM in general has scalability problem
ABC	• Efficiently used for multivariable, multimodal function optimization, simplicity, flexibility and robustness.	• Higher number of objective function evaluation • Require new fitness tests on new algorithm parameters.
AIS	• Has ability to recognize and to respond to different patterns and has an ability to learn, memorize, encode and decode information	• Limited ability to extrapolation
GA	• GAs do not break easily even if the inputs changed slightly, or in the presence of reasonable noise. • In searching a multi-modal state-space, large state-space, or n-dimensional surface, a genetic algorithm may offer significant benefits over more typical search of optimization techniques	• Repeated fitness function evaluation for complex problems • Genetic algorithms do not scale well with complexity • In many problems, GAs have a tendency to converge towards local optima or even arbitrary points rather than the global optimum of the problem

population is a weighted linear combination of sparse trees and each tree in this combination can be considered as a gene. This model outperformed other models like standard GP and RBF network with lowest average MAPE value 1.5716%.

Table 3 shows the comparative study between the single load forecasting techniques.

VI. HYBRID MODELS WITH TWO METHODS
Single methods for load forecasting often comes with several types of disadvantages including computing efficiency,

computational complexity, and high error percentage. Over the years, researchers have been working on building hybrid load forecasting methods and models to get better accuracy with minimum error rate. Hybrid models are generally a combination of two or more single methods where each method helps to make the forecasting more accurate and efficient. The single methods in a hybrid model is chosen according to its best advantages to contribute to the forecasting. SVM and ANN are the two most popular methods that are being hybridized with other single methods to gain the best

load forecasting model with minimal error percentage. SVM works well with unstructured and semi structured data to gain the best output. Kernel trick is the real strength of SVM's model where an appropriate kernel function can solve any complex problem. When the optimality is rounded, SVM can produce a unique solution. This forms a fundamental difference of SVM and Neural Networks, which produce multiple solutions based on local minima, which makes them not trustable over different samples. On the other hand ANN can learn and generalize from training data so there is no need for vast feats of programming. Like the "graceful degradation" found in biological systems, ANNs are particularly fault tolerant. Overall, there adaptive learning capability based on the data given for training made ANN one of the most popular method for load forecasting. This section reviews and shows comparison between the hybrid models developed based on SVM and ANN to maximize the best forecasting result. Also hybrid models with more than two single methods are discussed here with comparison between them.

A. SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine is one of the popular machine learning algorithms which was initially developed by Cortes and Vapnik [93] in 1995. Support Vector Regression (SVR) is a methodology to train the SVM where a function is estimated using observed data and is used for many machine learning tasks including time series prediction, regression analysis etc. In a given training data set:

$$\{(X_1, Y_1), \dots, (X_i, Y_i), \dots, (X_1, Y_1)\} \subset R_n \times R \quad (10)$$

In SVM for linear regression (SVR), the following function is to be estimated,

$$f(x) = <w, x> +b \quad w, x \in R_n, \ b \in R \quad (11)$$

Here the regularized risk function R is being minimized,

$$R(w, b) = C\frac{1}{l}\sum_{i=1}^{n} L(Y_i, f(X_i)) + \frac{1}{2}\|W\|^2 \quad (12)$$

where, the former term indicates the empirical error measured by loss function L. C is called the regularization constant which determined tolerated deviations from the loss function. The latter term consists of a regularizer W, which will make the function as flat as possible to control the function capacity.

SVM has been a popular choice among the machine learning enthusiasts since 1990. SVM is defined by a convex optimization problem (no local minima) for which there are efficient methods (e.g. Sequential Minimal Optimization). Besides SVMs deliver a unique solution, since the optimality problem is convex. This is an advantage compared to NNs, which have multiple solutions associated with local minima. Computationally expensive and slow running is one of the common drawbacks of SVM. To obtain the global minimum for load forecasting, SVM is an assuring learning tool. By providing a necessary parameter for the user-selected kernel function, this unique method can be optimized with the different type of kernel function and all the parameters of
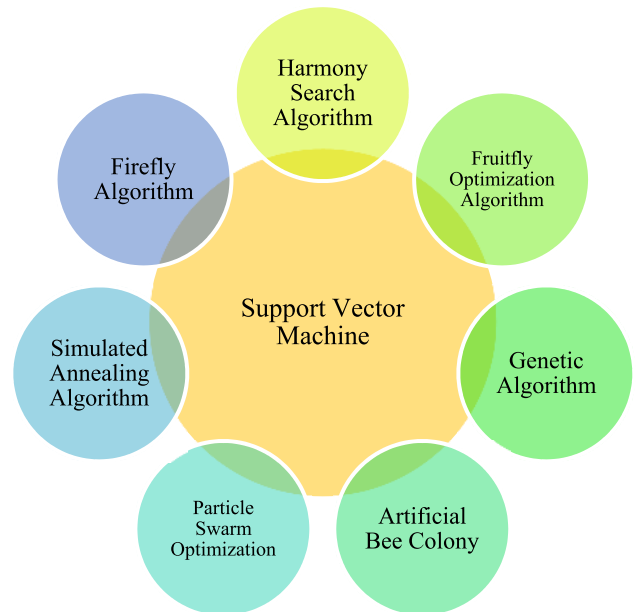
FIGURE 12. Pictorial representation of notable hybrid models with two methods based on support vector machine that are explained in the subsection.

SVM. Some of the most popular hybrid methods integrating SVM with other single methods are:

1. SVM and Broyden-Fletcher-Goldfarb-Shanno Firefly Algorithm
2. SVM and Harmony Search Algorithm
3. SVM and Fruit Fly Optimization Algorithm
4. SVM and Genetic Algorithm
5. SVM and Particle Swarm Optimization
6. SVM and Artificial Bee Colony
7. SVM and Simulated Annealing Algorithm

Fig. 12 presents the commonly used hybrid two-method models comprising of SVM. These methods are presented in the following subsection with their uses in load forecasting technology, the performance of which are further compared on the basis of their accuracy.

1) SVM AND BROYDEN–FLETCHER–GOLDFARB–SHANNO FIREFLY ALGORITHM (SVM-BFGSFA)

Firefly Algorithm (FA) is one of the recent swarm intelligence methods developed by Yang [94] in 2008 and is a kind of stochastic, nature-inspired, meta-heuristic algorithm that can be applied for solving the hardest optimization problems including load forecasting. Its unique ability to search and find the global and local optima concurrently made it a very popular algorithm for forecasting. This algorithm is based on the characteristics of the flashing light of fireflies. The flashing light produced by the fireflies by biochemical process of bioluminescence which helps to characterize them. Kavousi-Fard *et al.* [95] combined a Modified Firefly Algorithm (MFA) and the SVR model to enhance both the search ability and the convergence of FA. As a case study,

this work used half-hourly electrical power consumption data of New South Wales, the state of Victoria, and the State of Queensland in Australia to validate the predictability of the proposed combined model. To overcome the weaknesses of FA which are slow convergence speed and the imprecise accuracy of convergence, Xiao *et al.* [96] proposed a new combined model where they integrated multiple seasonal patterns, several neural networks, non-positive constraint theory and Broyden–Fletcher–Goldfarb–Shanno Firefly Algorithm (BFGSFA) where it has shown great result compared to the other single models. To retain the advantages of FA, such as powerful global exploration and exploitation abilities, and overcome the weakness of FA during the latter period of optimization, such as the slow convergence speed and the imprecise accuracy of convergence. BFGS is used when FA updates solutions in an iteration to find a local optimal solution to enhance the local optimization ability and the speed of the local convergence of the whole algorithm. This combined model can be useful to forecast electric load, power scheduling and can avoid power grid collapse and reduction of the spinning reserve capacity of thermal power plants. The average MAPE values of the combined model were 0.7138%, 1.0281%, 4.8394%, 0.9239%, 9.6316% and 7.3367% lower than other models e.g. Back-Propagation Neural Network, GA based ANN, Wavelet based NN, GRNN, ARIMA and Random Walk.

2) SVM AND HARMONY SEARCH ALGORITHM (SVM-HS)

Harmony Search (HS) is a meta-heuristic optimization search algorithm inspired by a musician effort to search for the better harmony [97]. It is called during the training phase to select the optimum parameter. In order to develop the process of determining the parameters of STLF model harmony search was adopted to LS-SVM by Zeng *et al.* [98] Their results showed the improvement in solution quality and a higher training speed. Interestingly, the MAPE value of the proposed methods is 0.77% lower than PSO while the training time and speed are 36.7% and 2.59% higher, respectively. From their study, it is being concluded that through the analysis of load influencing factors with GRNN, the accuracy of the forecasting model (HS-LS-SVM) can be improved. The key factors were selected at first by GRNN as the economic and non-economic factors affect the STLF. The proposed method achieves higher precisions and faster speed than BPNN, LS-SVM and PSO, and its correctness and efficacy are also verified.

3) SVM AND FRUIT FLY OPTIMIZATION ALGORITHM (SVM-FOA)

Pan [99] proposed Fruit Fly Optimization Algorithm (FOA) at first, based on food process of the fruit fly. The shorter program code made it popular compared to other optimization algorithm. FOA algorithm has been applied to various forecasting applications including load forecasting, as it can reach the global optimal solution fast. Li *et al.* [100] have proposed a hybrid model SVM-FOA in their work. Compared

to other heuristic optimization algorithm such as GA and Stimulated Annealing (SA), SVM-FOA, showed great results by minimum searching time to locate the global optimum including decent forecasting error (∼3%) for annual load. In another work, Cao and Wu [101] proposed a novel hybrid approach to forecast monthly electricity consumption, combining SVM with FOA. In this study, monthly electrical load consumption of China has been used since January, 2010 to December, 2015. This 72 months of dataset has been divided into training set (the preceding 60 months) and the testing set (last 12 months). The results shows that FOA has outperformed PSO algorithm in some cases.

4) SVM AND GENETIC ALGORITM (SVM-GA)

GAs are well suited to the concurrent manipulation of models with varying resolutions and structures since they can search non-linear solution spaces without requiring gradient information or a priori knowledge of model characteristics [102]. It provides an optimization solution to find a computer program of unspecified size and shape to solve, or approximately, a problem. GA has the ability to search through a space to find the nearly optimal solution. Fitness function is evaluated to measure the computer solving speed of a given problem [103]. One of the advantages of GA is that, it search parallel from a population of points, therefore, it has the ability to avoid being trapped in local optimal solution like traditional methods e.g. regression. It uses probabilistic selection rule rather than deterministic ones. GA works on the chromosome, which is encoded version of potential solutions' parameters, rather the parameters themselves. It uses fitness score which is obtained from objective functions, without other derivative or auxiliary information. In his work, Lee *et al.* [104] has shown that Genetic Programming (GP) model outperforms the results of regression methods to solve the LTLF problem.

GA has been hybridized with SVM to optimize the SVR parameter values in many studies. Hsu *et al.* [105], at first proposed a GA-SVR model to overcome the problem of SVR parameters. SVR has been integrated with an improved adaptive genetic algorithm (IAGA) [106] to optimize the ratio values of meteorological factors and electricity cost, outperforming state-of-the-arts. A separate hybrid genetic-based SVR model (HGASVR) has been proposed by Wu *et al.* [102]. The dataset used in this study was downloaded from EUNITE network. It consists of all-half-hour electricity values and average temperature value for 1997 and 1998. Additionally, the holidays of these two years also included in the dataset. To achieve better forecasting accuracy their proposed model can automatically optimize the SVR parameter integrating the real-valued genetic algorithm and inter genetic algorithm. In their study, it is proposed that the Poly kernel function could be an appropriate choice of SVR Kernel function in forecasting daily electricity load. The Poly kernel function might outperform the RBF kernel function in a non-linear electricity load forecasting problem. The optimal MAPE, RMSE and maximum error by HGASVR are 0.76, 7.73 and 20.88 respectively.

TABLE 4. Comparative study of SVM based hybrid methods.

METHODS	ADVANTAGE	DISADVANTAGE	COMPUTATIONAL COMPLEXITY	DIFFICULTIES
SVR-FA	Minimum number of average value iteration	May converge with the most optimal solution within a limited time because of its complexity	Reducing the local optima trapping possibilities	Optimal parameter selection
SVM-HS	High training speed and prediction accuracy	Complicated architecture	To maintain convergence speed	Optimal population size detection
SVM-FOA	Can reach global optimum solution fast	Complicated Architecture	To select the appropriate values of three parameters of the SVR model	Parameter determination
SVM-GA	Can identify the optimal values with lower prediction errors	Time consuming searching algorithm	To optimize all the parameters of SVR simultaneously	Optimizing the free parameters of SVM
SVM-PSO	Memory storage option/capability	Cannot overcome the local minimization drawback efficiently	Selecting the optimal parameters	Prone to premature convergence
SVM-ABC	Recurrent mechanism with superior capability to capture more data pattern information from past electric load data	Slow in sequential process	Lack of global search ability	Slow convergence speed
SVM-SA	Capable of non-linear mapping	Do not perform minimization training error	To select the free parameter	Parameter Determination

5) SVM AND PARTICLE SWARM OPTIMIZATION (SVM-PSO)

PSO was initially developed by Kennedy and Eberhart in 1995 [107]. It is similar to GA in that sense that the system is initialized with a population of random solutions [108]. It has a unique feature of storing all the good solutions of all the particles. For non-linear problems, a few parameters can be adjusted. Unlike GAs and SA, it has a memory storage function. Hong [109] proposed a hybrid SVM-PSO algorithm to locate optimal parameters of SVR model which can be used for load forecasting. It shows better result than SVM-GA and SVM-SA. One of the biggest drawbacks of PSO is that unlike GA and SA, it cannot overcome the local minimization drawback efficiently. Jiang *et al.* [110] designed an SVR based forecaster with a two-step hybrid parameter optimization scheme with Grid Traverse Algorithm for global to local parameterization, and with PSO for determining the best local parameters. Their method showed better performance compared to ARIMA (11.21%), SVM-GA (5.27%) and ANN (6.62%) with MAPE value of 2.53%.

6) SVM AND ARTIFICIAL BEE COLONY (SVM-ABC)

ABC was applied for searching the global minimum of three well-known test functions (Sphere function, Rosenbrock valley and Rastrigin function) [74]. It can make a good balance between global and local searches by conducting both in each iteration instead of initiating global search at the beginning and the local search at the end stage of PSO [111]. For better parameter determination, Duat *et al.* [112] proposed a modified ABC algorithm which can provide a better accuracy of an electric load forecasting model based on Least Square Support Vector Machine (LS-SVM).

7) SVM AND SIMULATED ANNEALING ALGORITHM (SVM-SA)

Simulated annealing algorithm is an optimization technique, analogous to the annealing process of material physics. SVMs have been employed to solve nonlinear regression and time series problems. Pai and Hong [113] applied SA algorithm to choose the parameters of a SVM model. The SVM-SA model performs structural risk minimization rather than minimizing the training errors. This method outperforms the ARIMA and GRNN model in generalized performance. Wang *et al.* [114] proposed a new optimal model Stimulated Annealing Particle Swarm Optimization Algorithm that combines the advantages of PSO and SA algorithm. The model is able to enhance the accuracy and improved the convergence ability and reduced operation time by numerical experiment.

Table 4 presents a comparative study of the advantages and disadvantages between the SVM based hybrid methods. A comparative analysis of the MAPE of these methods are presented in Table 5. Also the value of these metrics are different for different datasets and parameters. Therefore it is quite difficult to compare the results of different techniques. Also, there is no such task where all the methods are experimented in a single dataset, to find the comparison between them. In this research work, the best accuracy of each discussed method are tabulated in the subsequent sections from different forecasting methods.

B. ARTIFICIAL NEURAL NETWORK (ANN)

ANNs are processing devices (algorithms or actual hardware) that are loosely modeled after the neuronal structure of the mammalian cerebral cortex but on much smaller scales. A large ANN might have hundreds or thousands of processor

TABLE 5. Comparative MAPE, RMSE values and factors of SVM based hybrid model.

Methods	MAPE	Factors	Reference Paper
SVR-FA	1.8051	N/A	[95]
SVM-HS	4.579	N/A	[115]
SVM-FOA (GRNN)	3.679	N/A	[101]
SVM-GA	0.76	Average Daily Temperature	[102]
SVM-PSO	1.92	Daily min. & max. temp, avg. temp, relative humidity, wind force	[116]
SVM-ABC	0.5268	Real time temp., weather	[117]
SVM-SA	1.76	Temperature	[113]

units, whereas a mammalian brain has billions of neurons. ANNs provide an analytical alternative to conventional techniques which are often limited by strict assumptions of normality, linearity, variable independence etc. Because an ANN can capture many kinds of relationships it allows the user to quickly and relatively easily model phenomena which otherwise may have been very difficult or impossible to explain otherwise. ANN is popular choice for load forecasting technique. A recurrent type ANN, proposed by Baek [118], have shown an MAPE value of 1.57% for MTLF in South Korea. In this study, dataset includes daily load consumption, real-time temperature, weather, and day type from July, 2011, a week before 168h load data as the training set and July 8th, 24 hour of the same year as test set. The power consumption value has been taken in 15 minutes interval same as temperature record. Whereas, the day and weather type recorded once a day.

MLP, another architecture of ANN model was employed by Park *et al.* [119] to STLF. One of the major disadvantage of this method is it takes large MLP structure to train the real data and it also raise redundancy issues. Some of the most popular hybrid methods integrating ANN with other single methods are:

1. ANN and Fruit Fly Optimization Algorithm (ANN-FOA)
2. ANN and Firefly Algorithm (ANN-FA)
3. ANN and Clustering Techniques (ANN-CT)
4. ANN and Neural Fuzzy Interference System (ANN-NFIS)
5. ANN and Artificial Immune System (ANN-AIS)
6. ANN and Wavelet Transform (ANN-WT)
7. ANN and Particle Swarm Optimization (ANN-PSO)
8. ANN and Genetic Algorithm (ANN-GA).

Fig. 13 presents the commonly used hybrid two-method models comprising of ANN. These methods are presented in the following subsection with their uses in load forecasting technology, the performance of which are further compared on the basis of their accuracy.
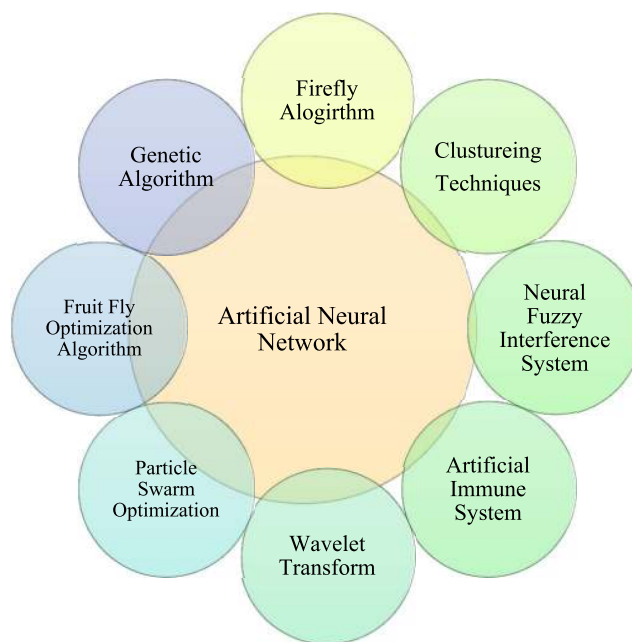
FIGURE 13. Pictorial representation of notable hybrid models with two methods based on artificial neural network that are explained in the subsection.

1) ANN AND FRUIT FLY OPTIMIZATION ALGORITHM (ANN-FOA)

Pan *et al.* [99] discussed about the basic fundamental part of the FOA algorithms which helps to find the maximal value and minimal value. The searching viability of FOA algorithm is a help for parameter selection of neural network. Li *et al.* [120] combined GRNN and FOA for annual power load forecasting where FOA is used to determine the spread parameter value of the GRNN model. The MAPE and MSE values of their proposed model were 1.149% and 1.421 respectively which is also compared with the other five forecasting models namely FOA-GRNN, GRNN, PSO-GRNN, Sparse Approximation Scheme for Least Square SVM, and Ordinary Least Squares-Likelihood Ratio. It was found that FOA-GRNN shows the best performance compared to other methods.

2) ANN AND FIREFLY ALGORITHM (ANN-FA)

In ANN-FA method, FA is used to create the nonlinear mapping and ANN works to achieve the learning ability. It helps to develop relatively efficient and accurate forecasting model, but shows a poor performance with respect to RMSE. Kavousi-Fard *et al.* [121] combined FA and ANN to develop a more efficient and accurate forecasting model where FA was used to create the nonlinear mapping and ANN works to achieve the learning ability. Their proposed hybrid MFA and ANN model mathematical expression is

$$X_k = [w_{i,1}, w_{i,2}, \ldots, w_{i,Mw}, b_{i,1}, b_{i,2}, \ldots, b_{i,Mb}]_{(1,M)}$$
$$M = M_w + M_b \qquad (13)$$

where, X_k is the randomly selected firefly from a set of fireflies, w_i is the adjusting coefficient for the ANN arranged in an M dimensional row vector, and M_w, M_b are the number of weighting and biasing factors, respectively. In addition, by comparing ANN-MFA method with the other methods showed that the proposed method results very low forecasting value of 2.0322 which is a more suitable result than the others. Another study was done by Xiao *et al.* [96] where he proposed a combined model of FA neural network where sufficient work is done to reduce the error of the load forecasting.

3) ANN AND CLUSTERING TECHNIQUE (ANN-CT)

Clustering Technique (CTs) was developed by Driver and Kroeber in 1932.The major advantage of CT is suitable for datasets with compact spherical clusters that are well-separated. Hernández *et al.* [122] discussed about the STLF using two different data set for microgrid where they amalgamated the K-means clustering algorithms with MLP. This model shows MAPE value for the data Set A 15.34% and data Set B 16.69%. Quilumba *et al.* [123] used the CT to group customers behavior based on similarities in consumption. The average MAPE value was reduced by 1.7%, increasing the number of clusters from one to four and showed a good forecasting accuracy. Fahiman *et al.* [124] combined the DL and clustering methods to improve the accuracy of load forecasting where clustering is used to identify the customers behavior into sub-populations based on similar demand profiles. This model used K- clusters $C = \{c_1, c_1, \ldots, c_k\}$ to train a neural network with k prediction models $F_C = \{F_{c_1}, F_{c_2}, \ldots, F_{c_k}\}$ as:

$$F_{aggr} = \sum_{i=1}^{k} F_{c_i} \qquad (14)$$

For real-time forecasting, the DNN based K-clusters forecasting model shows the best accuracy among the four proposed methods discussed by the author. The author suggest to use dynamic clustering methods for real-time forecasting.

4) ANN AND NEURAL FUZZY INFERENCE SYSTEM (ANN-NFIS)

This forecasting method is a combination of ANN with the neural fuzzy inference system (NFIS). ANN is a method to obtain an approximate solution through Artificial Intelligence techniques, which is based on the imitation of the functioning of brain [125]. ANNs are organized as layers, i.e. input layer, hidden layer and output layer. Fuzzy inference use the FL to process the mapping from given input variable to an output variable. Using fuzzy logic (IF-THEN) \ rules, the variables are matched and the response is acquired through fuzzy implications [126]. The researchers combine both the techniques to enrich the performance by controlling decision-making and analyzing data. To find the best parameters, fuzzy classification technique can help the neural network by grouping the rule base. This is extremely helpful to forecast with irregular datasets which is usual in most of the cases of forecasting. Neuro-fuzzy modeling is also easy to incorporate human expertise about the target system directly into the modeling

process [127], [128]. Khotanzad *et al.* [129] combined FL along with ANNs to develop a two-stage fuzzy logic where it is found that the MAPE is less than the Artificial Neural Network based Short Term Load Forecaster.

5) ANN AND ARTIFICIAL IMMUNE SYSTEM (ANN-AIS)

The advantage of the immune system is that it has the powerful information processing capabilities which made the system highly parallel. This system is inspired by the immunology of human body. Immune Algorithm (IA) was applied by Yong *et al.* [89] to design back propagation neural network (BPNN). It was named as Artificial Immune Network (AIN). This MAPE value of this model is 2.52% whereas the AIN MAPE is 2.038%. A new model was developed by Hamid and Rahman [130] where the ANN is trained by the AIS. It is capable to provide a comparable forecast of ANN with BP.

6) ANN AND WAVELET TRANSFORM (ANN-WT)

Wavelet is a complement to classical Fourier decomposition method. Wavelet transforms allow the components of a non-stationary signal to be analyzed and the filters to be constructed for stationary and non-stationary signals. To decompose the loads into multiple frequency components WT is prescribed in various works. Jawerth and Sweldens [131] focused on multi-resolution analysis of WT. Bashir and El-Hawary [132] used the back propagation algorithm for solving short-term load forecasting and build a structure of wavelet neural network. This technique showed the average percentage error is less than the individual ANN. Zhang and Dong [133] included Artificial Neural models along with wavelet transformation to show that the useful information can be captured on various time scales. The proficiency of this method has been tested by the Australian electricity market to predict electricity demands in short term data series and showed promising results. Guan *et al.* [134] integrated wavelet neural networks (WNN) with data pre-filtering. Li *et al.* [135] utilized wavelet decomposition to reduce the non-stationary load sequence and performed an augmented Dickey-Fuller test to determine the stationary components of the decomposition. The forecasting of the wavelet features was analyzed using a second-order gray neural network for performing STLF with MAPE value of 2.41% for 9 layers. Wavelet based decomposition are used for the construction of probabilistic load forecasting models, combined with tree based Random Forest and quantile regression forests as proposed by Alfieri and De Falco [136]. El-Hendawi and Wang [137] forecasted the load of Ontario, Canada and successfully reduced 20% of MAPE value with the ensemble method of ANN-WT in comparison to conventional NN.

7) ANN AND PARTICLE SWARM OPTIMIZATION (ANN-PSO)

El-Telbany and El-Karmi [138] combined the PSO with ANN where PSO used to train the feed forward network for load forecasting of the Jordanian electricity system. The authors found that PSO algorithm performed better than the BP algo-

TABLE 6. Comparative study of ANN based hybrid methods.

METHODS	ADVANTAGE	DISADVANTAGE	COMPUTATIONAL COMPLEXITIES	DIFFICULTIES
ANN-FOA	Better searching viability for parameter selection of Neural Network	Unstable search route and slower convergence speed	Tuning the spread parameter value of the GRNN model	Simply trapped in local optimum value
ANN-FA	Can solve the local optimal drawback of neural network	Requires large number of iteration and proper parameter setting	Slow speed of convergence and the imprecise accuracy of convergence	To optimize the weight coefficients
ANN-CT	Can identify and model non-linear input/output complex relationship	Large number of samples and clusters required	Selection of input variables	Selecting optimal number of clusters, lack of class levels
ANN-NFIS	Can forecast irregular data set	Gradient descent slows down the training speed and increases the training time for complex task	Lower speed and longer run-time required	Lack of real time response
ANN-AIS	Fast searching ability	AIS has limited ability to extrapolation	Generalizing capability loss due to overspecialized leading	To maintain the proper hidden layer
ANN-WT	Loads can be decomposed into multiple frequency components	Depends on frequency resolution	To modify the wavelet network parameters	Shift invariance problem, Wrong frequency order
ANN-PSO	Computationally inexpensive and performs well in high load fluctuation	Requires many function evaluations to find optimal solution	Optimal number of swam selection	Degradation of ANN performance due to overfitting and memorization
ANN-GA	Convergence to local minima problem can be solved	Difficult to know which ANN parameters must be set by the GA for better performance	Time complexity for high number of iterations	Convergence is highly dependent on initial solution

rithm. Due to the over trained and complexity, BP (trained by Neural Network) showed the poor performance. Another application of ANN-PSO is made by Liu et al. [139]. They applied PSO to optimize the parameter of the ANN. The superiority of PSO is that from a set of possible alternatives it can find best element and also computationally inexpensive, easily implemented, and does not require gradient information of the objective function but only its values. Zhang and Ma [140] integrated the PSO algorithm and the RBF neural network algorithm where PSO used to optimize the weights and RBF neural network to learn the accuracy. This model showed better performance the typical RBF network.

8) ANN AND GENETIC ALGORITHM (ANN-GA)
Ling et al. [141] proposed a novel neural network based on GA algorithm. The proposed network's parameter can be tined by a GA with arithmetic crossover and non-uniform mutation. This model has a two activation function namely static activation function (SAF) and dynamic activation function (DAF). If v_{ij} is the synaptic connection weight from the i-th input node x_i to the j-th neuron, n_h is the nodes in hidden layer, $net_s^j(\cdot)$ is the j-th SAF, $net_d^j(\cdot)$ is the j-th DAF, and $net_o^l(\cdot)$ is the activation function for l output neurons, m_d^j and σ_d^j are the dynamic mean and dynamic standard deviation for the j-th DAF, their proposed model mathematical expression for the daily electric load forecasting is:

$$y_l(t) = net_0^l \left(\sum_{j=1}^{n_h} net_d^j (net_s^j \left(\sum_{i=1}^{24} x_i v_{ij} \right), m_d^j, \sigma_d^j) w_{jl} \right)$$
$$l = 1, 2, \ldots, 24. \quad (15)$$

This model showed the MAPE value is less than the traditional neural model, also providing the training information which is better on proposed model. An integration study was performed by Azadeh et al. [142] to developed a logarithmic-linear model which was later studied on the Iranian agriculture sector. The author proposed ANN model MAPE value which is much better than time series model. Ventura et al. [143] studied on which parameters are best for the implement the integration between GA and ANN. Previously mentioned PSO and ANFIS has also been integrated with GA and Gaussian process regression to optimize the input of the ANFIS based PV power forecasting for the plants [144].

Table 6 presents the comparative study of the advantages and disadvantages, computational complexities between the ANN based hybrid methods. A comparative analysis of the MAPE factors of the ANN based hybrid methods are presented in Table 7.

VII. HYBRID MODELS WITH MORE THAN TWO METHODS
In recent times, scientist and researchers are working more on developing hybrid models with more than two single methods where each single method has their own purpose of work. Some of these hybrid models are discussed below:

A. GENERALIZED NEURAL NETWORK-WAVELET TRANSFORM-GENETIC ALGORITHM WITH FUZZY LOGIC (GNN-WT-GAF)
To overcome the major narrowness of single and double hybrid methods, Chaturvedi et al. [150] proposed a technique where Generalized Neural Network (GNN) integrates

TABLE 7. Comparative MAPE, MSE/RMSE values and factors of ANN based hybrid methods.

Methods	MAPE	Factors	Reference Paper
ANN-FOA	1.149	Temperature	[120]
ANN-FA	1.1808	N/A	[96]
ANN-CT	0.4935	N/A	[145]
ANN-NFIS	0.000396	Temperature, wind speed, humidity	[146]
ANN-AIS	2.457	Temperature	[130]
ANN-WT	1.111	Temperature, Weather condition (sunny, cloudy, rainy etc.)	[147]
ANN-PSO	1.9882	Temperature, wind speed, humidity	[148]
ANN-GA	0.020	Temperature, humidity	[149]

with WT to train the genetic algorithm with fuzzy system (GAF) namely GNN-WT-GAF model. This proposed model is useful to limiting the drawbacks of ANN, to improve the convergence of GA and enhance the performance of fuzzy concepts. This proposed method has been tested using datasets from a 15 MVA, 33/11 kV substation at Dayalbagh Educational Institute (D.E.I.), Agra, India where datasets are divided into four wavelet components where wavelet is a mathematical function used to divide a given function of continuous- time signal into different scale components, namely one approximate component (a_3) and three detailed components (d_1, d_2, d_3).These component are used to train the GNN model with given load pattern at time t, t_1, t_2 as input and $t + 1$ as output. In final stage the wavelet components of GNN was trained by an adaptive genetic algorithm using fuzzy system (GAF). Furthermore, the proposed GNN-WT-GAF techniques compare with the other methods such as ANN-BP, GNN-BP, GNN-GAF, GNN-WT-BP and GNN-WT-GAF. This proposed method shows a RMSE value 0.0486 kW which is a more suitable result than the others.

B. EMPIRICAL MODE DECOMPOSITION-PARTICLE SWARM OPTIMIZATION-SUPPORT VECTOR REGRESSION (EMD-PSO-SVR) OR ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (EMD-PSO-ANFIS)

For enhance the accuracy of load forecasting, EMD-PSO are garnering popularity due to the ability of EMD to decompose complicated load data into different intrinsic mode functions (IMFs), which can further be optimized by PSO. Wang and Wang [151] developed a load forecasting method where EMD-PSO is combined with SVR. Semero *et al.* [152] proposed a similar approach by integrating EMD-PSO with ANFIS. Their models have three superior capability. Firstly

load data is divided into a number of IMFs components and one residue by EMD, secondly prognosis IMFs and residual value separately by SVR or ANFIS, thirdly PSO select the parameter of SVR or ANFIS automatically. Then the combination is used to make aggregate calculation and predict the result. For testing the ability of the proposed model they used the actual daily peak load data of State Grid Handan Electric Power Company in China for SVR and of a microgrid in Beijing for ANFIS. The proposed EMD-PSO-SVR model is compared with the various model such as EMD-SVR, PSO-SVR, SVR which showed that the proposed method results very low error values, i.e. MAPE value of 2.7510, a RMSE of 0.0595, and an MAE of 0.0414 for an alternate case study.

C. ENSEMBLE EMPIRICAL MODE DECOMPOSITION-EXTREME LEARNING MACHINE-GRASSHOPPER OPTIMIZATION ALGORITHM (EEMD-GLM-GOA)

Wu *et al.* [153] combined three single method for short term load forecasting to propose a new excellent hybrid model,which is assembled by the Ensemble Empirical Mode Decomposition (EEMD), ELM and grasshopper optimization algorithm (GOA). This model is superior to find the suitable weight and threshold values of ELM by GOA. This model has been examined using datasets from five main states in Australia, New South Wales, Queensland, Tasmania, South Australia and Victoria. In training phase, original electrical data set was decomposed by EEMD as well as suitable parameter are select by GOA for ELM. Then a non-linear model was constructed by ELM. The propose EEMD-ELM-GOA model was compared with three other types of models, i.e. Ensemble Empirical Mode Decomposition – Extreme Learning Machine – Dragonfly Algorithm (EEMD-ELM-DA), Ensemble Empirical Mode Decomposition – Extreme Learning

Machine – Grey Wolf Optimizer (EEMD-ELM-GWO), and EEMD-ELM-PSO where proposed model has the best performance.

D. GLOBAL HARMONY SEARCH ALGOTRITHM-FUZZY TIME SERIES-LEAST SQUARES-SUPPORT VECTOR MACHINE (GHSA-FTS-LS-SVM)

Chen *et al.* [115] proposed a hybrid GHSA-FTS-LS-SVM electrical load forecasting model based on Global Harmony Search Algorithm (GHSA) with Least Squares Support Vector Machines (LS-SVM) and the Fuzzy Time Series (FTS). In this proposed model, FTS calculate the clustering center of each cluster where the LS-SVM is used to model the resultant series, which is optimized by GHSA. Datasets from the Guangdong Province Industrial Development was used to test the accuracy of the proposed model result, which was also compared with different types of models such as ARIMA and other algorithms hybridized with LS-SVM including PSO, HS and GA. In addition, the proposed GHSA-FTS-LS-SVM techniques shows MAPE, MAE, and RMSE values of 3.709, 14.358 and 18.180 respectively which has a more suitable result than the others.

E. T-COPULA-IMPROVED EMPIRICAL MODE DECOMPOSITION-DEEP BELIEF NETWORK (T-COPULA-IEMD-DBF)

The proposed hybrid technique by Haq and Ni [14] is designed for STLF, which are tested by the real time data from Australia and the United States of America. Dataset included weather data, time categorical data, social data and energy load demand for particular sampling time. Load demand time series, after decomposing into low frequency components with improved Empirical Mode Decomposition (IEMD), the effects of the loss components are retrieved using T-Copula based correlation analysis. On deriving the peak load indicative binary variables from value at risk (VaR), the combined data is fed to DBN for STLF. For five locations in Australia, the proposed method is compared to the standard NN, DBN and EMD-DBN data derived from [154] and for the standard NN, DBN and Copula-DBN derived from Texas, USA [27]. The model outperformed with respect to MAPE and RMSE for most of the cases. On average, the values for the hybrid models are reduced by 21.19% and 16.93% in case of Australia, and by 15.27% and 13.86% in case of the USA.

F. GENETIC ALGORITHM-NON LINEAR AUTO REGRESSIVE WITH EXOGENOUS INPUT-NEURAL NETWORK (GA-NARX-NN)

Jawad *et al.* [155] combined GA based training method for NARX-NN for 168 hours ahead STLF and 1 month ahead MTLF problems. A wind speed pattern recognition was also designed with the model. Preliminary model was designed using elitist GA, an advanced form of GA to improve the convergence in real-number coding. The chromosome population of GA are provided to NARX-NN as the initial weight to assess the fitness function for iteration. The

concluding performance of the proposed model was compared with different single methods including auto-regressive (AR) with exogenous input (ARX), AR moving average with exogenous input (ARMAX), NARX, regression tree (RT), Levenberg-Marqardt (LM)-ANN and LM-NARX-NN. Experimentation showed a regression value greater than 0.99 for load forecasting, showing excellent values of 1.12% MAPE and 1.39% RMSE (with 0.00036 variance) tested for MTLF from ERCOT data for the month of December.

G. FUZZY COMBINATION WEIGHT-EMPIRICAL MODE DECOMPOSITION AND KALMAN FILTERING-BAT ALGORITHM-SUPPORT VECTOR MACHINE (FCW-EMD AND KF-BA-SVM)

A multiple method based hybrid model is proposed by Liu *et al.* [156] which addresses multiple decomposition based factors into consideration (such as preprocessing, same day selection, sequence decomposition, selection and optimization). Fuzzy Combination Weight (FCW) assists the selection of dense days, reducing the data for forecasting. After EMD decomposes the components into series IMF, Bat Algorithm (BA) is then used to optimize the SVM parameters, the forecasted values of which are fine-tuned by Kalman Filter (KF). SVM is in the core of the complete model. The model is tested with the transformer substation data in South China for specific days with varying load profile. The model excelled over KF-BA-SVM, EMD and KF-BA-SVM, showing MAPE value of 1.10, which was 1.77 and 0.96 less than those methods respectively. In comparison with GA-SVM, PSO-SVM and EMD-SVM, where these models showed MAPE value of 13.5115%, 12.932% and 4.2519% respectively for a holiday of 2015, the proposed combination showed 1.9052%, outperforming all the aforementioned models.

H. GENETIC ALGORITHM-PARTICLE SWARM OPTIMIZATION-ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (GA-PSO-ANFIS)

Semero *et al.* [157] proposed a hybrid model for VSTLF in microgrids integrating GA, PSO, and ANFIS. The binary genetic algorithm selects important predictors that notably influence the load pattern among a number of input variables whereas PSO algorithm is used to optimize an ANFIS-based model for VSTLF. The proposed model was compared with other three models: BP neural network, ARIMA, and persistence models. The proposed method shows 40% reduced average execution time for one-step-ahead load forecasting. Apart from these notable hybrid methods, novel combined and complex algorithms such as Bidirectional Gated Recurrent Unit (Bi-GRU) based EGA-STLF [158], two layer hybrid neural network and three stage based enhanced ELITE (E-ELITE) framework and consensus-based mixed integer PSO assisted TRUST-TECH (CMPSOATT) method [159], EMD and fuzzy RBF-AR method for energy time series forecasting [160], and multiple scale Variational Mode Decomposition (VMD), K-Means Clustering and LSTM

TABLE 8. Comparative study of hybrid methods (more than two model).

METHODS	ADVANTAGE	DISADVANTAGE	COMPUTATIONAL COMPLEXITIES	DIFFICULTIES
GNN-WT-GAF	Have the ability to overcome the limitation of ANN and BP training algorithm	Requires manually change the best fitness over generations.	To maintain fitness value of chromosomes, optimal settings etc.	To keep the constant crossover and mutation probabilities.
EMD-PSO-SVR/ANFIS	Ability to decompose load data into IMF components.	Lack of ability to select the suitable three parameter, all the results are not integrated together as the final prediction for SVR	Stack in global optimization.	Dealing with nonlinear and non-stationary data
EEMD-GLM-GOA	Ability to search suitable parameter of threshold and weight values, need less time.	Unable to predict abnormal events	To select the traditional parameters randomly	To minimize the parameter selection error
GHSA-FTS-LS-SVM	Improving the computing accuracy and speed through a novel category of bionic algorithm, higher prediction accuracy and better computational speed	Parameter needs to select manually	Selection of regularization parameter ζ and the RBF kernel function parameter μ	Optimization of the parameter
T-COPULA-IEMD-DBN	Peak load indicative variable for VaR by T-Copula enhances peak time forecasting accuracy, aided by near-perfect fitting of nonlinear component by DBN	Repetition of process due to stochastic nature of exogenous variables	Determination of extreme points of observation interval, upper tail dependence parameter	Capture long-term dependencies for household data [164]
GA-NARX-NN	Use of elitist GA for fast convergence, less complexity and can be used for both STLF and MTLF	Manual choice for GA parameters	Length of chromosomes change with the change in the number of hidden neurons	Data complexity and data size reduce performance
FCW-EMD AND KF-BA-SVM	Integrates key processes, novel weight assignment theory and improved optimizer and selector algorithm predict with more accuracy	Inability to learn the temporal features from input data	Selection of subjective and objective weight	Proper parameterization of different methods

assisted for short term wind power forecasting [161] are being analyzed. These complex algorithms work for specific cases, and are being researched for general applications of load forecasting for different time horizons. Table 8 summarizes the advantages, disadvantages, computational complexity and difficulties in designing the notable models for load forecasting.

VIII. OUTCOMES
The major findings after performing a comparative study of different state-of-the-arts of predictive models are enlisted as follows:

- For designing a particular load forecaster, forecasting time, weather of the location and economy of the system have significant effect on the accuracy.
- Previous load data can contribute to the augmentation of the accuracy.
- The values from STLF can be approximated to MTLF or LTLF by adding econometric variables, load data or by performing statistical operations.
- The values of MAE, RMSE and MAPE can be utilized to evaluate the accuracy of any single or hybrid predictive model.
- Even though single predictive models can forecast loads with substantial precision, different models are integrated to improve the value of the forecast.

- Among the single methods mentioned, ELM provides fast training, FCM provides better error analysis index, FRBS has the universal approximation capability and MLP has the highest classification accuracy.
- SVM or ANN based double predictive models have shown improved results compared to the combination of other single predictive methods for load forecasting.
- More than two methods are designed in cascade to perform sequentially and improve the accuracy of each individual output.
- Each hybrid method, having their own merits, perform better than the combination of two methods as amalgamated methods contribute to the fine tuning of the parameters.

IX. CONCLUSION
As electrical load prediction requires precision and frequent check for the varying parameters, researchers have worked on numerous ML models to enhance the performance of the existing forecasting technologies. This paper has presented a comprehensive study of notable single methods and delineated the state-of-the-art for constructing hybrid models. Different load forecasting technologies with respect to time horizons have been described in the study, based on which various prediction models have been enlisted for performing an extensive literature study and comparative analysis. The combination of two or more predictive models

has been shown in the literature to construct hybrid models for demanded accuracy. SVM, ANN, and their relevant models have been proven fruitful, as these schemes have showcased excellent opportunities in achieving a well-organized power system utility where the demand load can be predicted with a minimum error percentage.

Statistical measurement has been utilized to compare the percentage error to assist in the selection of the most appropriate method for the particular forecasting routine. The outcome of the research has been enlisted to aid the readers in acquiring a detailed knowledge of contemporary predictive models, which can prove useful in selecting a particular model to design new hybrid forecasters as per the feature of the individual methods. As hybrid models with more than two methods have the considerable advantage compared to single or double models, future research will be focused on the design and comparison of hybrid models with three or more methods which will address the disadvantages of the current forecasters.

REFERENCES

[1] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proc. IEEE*, vol. 88, no. 2, pp. 163–169, Feb. 2000, doi: 10.1109/5.823996.

[2] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Feb. 2001, doi: 10.1109/59.910780.

[3] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation technique," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1788–1794, Nov. 1994, doi: 10.1109/59.331433.

[4] W. Gao, A. Darvishan, M. Toghani, M. Mohammadi, O. Abedinia, and N. Ghadimi, "Different states of multi-block based forecast engine for price and load prediction," *Int. J. Elect. Power Energy Syst.*, vol. 104, pp. 423–435, Jan. 2019, doi: 10.1016/j.ijepes.2018.07.014.

[5] D. Sáez, F. Ávila, D. Olivares, C. Cañizares, and L. Marín, "Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 548–556, Mar. 2015, doi: 10.1109/TSG.2014.2377178.

[6] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 306–316, May 2008, doi: 10.1109/TPWRS.2008.919318.

[7] Y. S. Manjili, R. Vega, and M. M. Jamshidi, "Data-analytic-based adaptive solar energy forecasting framework," *IEEE Syst. J.*, vol. 12, no. 1, pp. 285–296, Mar. 2018, doi: 10.1109/JSYST.2017.2769483.

[8] G. Li and H.-D. Chiang, "Toward cost-oriented forecasting of wind power generation," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2508–2517, Jul. 2018, doi: 10.1109/TSG.2016.2614341.

[9] S.-V. Oprea and A. Bara, "Machine learning algorithms for short-term load forecast in residential buildings using smart meters, sensors and big data solutions," *IEEE Access*, vol. 7, pp. 177874–177889, 2019, doi: 10.1109/ACCESS.2019.2958383.

[10] S. Ruzic, A. Vuckovic, and N. Nikolic, "Weather sensitive method for short term load forecasting in electric power utility of serbia," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1581–1586, Nov. 2003, doi: 10.1109/TPWRS.2003.811172.

[11] E. A. Feinberg and D. Genethliou, "Load forecasting," in *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence*, J. H. Chow, F. F. Wu, and J. Momoh, Eds. Boston, MA, USA: Springer, 2005, pp. 269–285.

[12] G. Singh, D. S. Chauhan, A. Chandel, D. Parashar, and G. Sharma, "Factor affecting elements and short term load forecasting based on multiple linear regression method," *Int. J. Eng. Res. Technol.*, vol. 3, no. 12, pp. 1–5, Dec. 2014.

[13] L. Katzir, "The effect of system characteristics on very-short-term load forecasting," *Przegl d Elektrotechniczny*, vol. 1, no. 11, pp. 121–125, Nov. 2015, doi: 10.15199/48.2015.11.31.

[14] M. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," *IEEE Access*, vol. 7, pp. 125413–125423, 2019, doi: 10.1109/ACCESS.2019.2937222.

[15] N. Abu-Shikhah, F. Elkarmi, and O. M. Aloquili, "Medium-term electric load forecasting using multivariable linear and non-linear regression," *Smart Grid Renew. Energy*, vol. 2, no. 2, pp. 126–135, 2011, doi: 10.4236/sgre.2011.22015.

[16] S. Ye, G. Zhu, and Z. Xiao, "Long term load forecasting and recommendations for China based on support vector regression," *Energy Power Eng.*, vol. 4, no. 5, pp. 380–385, 2012, doi: 10.4236/epe.2012.45050.

[17] S. Fallah, R. Deo, M. Shojafar, M. Conti, and S. Shamshirband, "Computational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions," *Energies*, vol. 11, no. 3, p. 596, Mar. 2018, doi: 10.3390/en11030596.

[18] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016, doi: 10.1016/j.ijforecast.2015.11.011.

[19] W. Charytoniuk and M.-S. Chen, "Very short-term load forecasting using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 15, no. 1, pp. 263–268, Feb. 2000, doi: 10.1109/59.852131.

[20] L. Ghods and M. Kalantar, "Different methods of long-term electric load demand forecasting a comprehensive review," *Electron. Eng.*, vol. 7, no. 4, pp. 249–259, 2011.

[21] R. Dechter, "Learning while searching in constraint-satisfaction problems," in *Proc. 5th AAAI Nat. Conf. Artif. Intell.*, Philadelphia, PA, USA, Aug. 1986, pp. 178–183. Accessed: Apr. 25, 2020.

[22] P. Kuang, W.-N. Cao, and Q. Wu, "Preview on structures and algorithms of deep learning," in *Proc. 11th Int. Comput. Conf. Wavelet Activ Media Technol. Inf. Process. (ICCWAMTIP)*, Dec. 2014, pp. 176–179, doi: 10.1109/ICCWAMTIP.2014.7073385.

[23] S. Ryu, J. Noh, and H. Kim, "Deep neural network based demand side short term load forecasting," *Energies*, vol. 10, no. 1, p. 3, 2016.

[24] Y. He, J. Deng, and H. Li, "Short-term power load forecasting with deep belief network and copula models," in *Proc. 9th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, vol. 1, Aug. 2017, pp. 191–194, doi: 10.1109/IHMSC.2017.50.

[25] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018, doi: 10.1109/TSG.2017.2686012.

[26] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018, doi: 10.1109/TPWRS.2017.2688178.

[27] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 2, pp. 127–136, Apr. 2019, doi: 10.1109/TETCI.2018.2880511.

[28] X. Tang, Y. Dai, Q. Liu, X. Dang, and J. Xu, "Application of bidirectional recurrent neural network combined with deep belief network in short-term load forecasting," *IEEE Access*, vol. 7, pp. 160660–160670, 2019, doi: 10.1109/ACCESS.2019.2950957.

[29] X. Tang, Y. Dai, T. Wang, and Y. Chen, "Short-term power load forecasting based on multi-layer bidirectional recurrent neural network," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 17, pp. 3847–3854, Sep. 2019, doi: 10.1049/iet-gtd.2018.6687.

[30] S. Motepe, A. N. Hasan, and R. Stopforth, "Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms," *IEEE Access*, vol. 7, pp. 82584–82598, 2019, doi: 10.1109/ACCESS.2019.2923796.

[31] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, pp. 88058–88071, 2019, doi: 10.1109/ACCESS.2019.2926131.

[32] X. Kong, C. Li, F. Zheng, and C. Wang, "Improved deep belief network for short-term load forecasting considering demand-side management," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1531–1538, Mar. 2020, doi: 10.1109/TPWRS.2019.2943972.

[33] L. Sehovac and K. Grolinger, "Deep learning for load forecasting: Sequence to sequence recurrent neural networks with attention," *IEEE Access*, vol. 8, pp. 36411–36426, 2020, doi: 10.1109/ACCESS.2020.2975738.

[34] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 470–482, Jan. 2020, doi: 10.1109/TSG.2019.2924183.

[35] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.

[36] Z.-H. You, S. Li, X. Gao, X. Luo, and Z. Ji, "Large-scale protein-protein interactions detection by integrating big biosensing data with computational model," *BioMed Res. Int.*, vol. 2014, Aug. 2014, Art. no. 598129. Accessed: Apr. 26, 2020. [Online]. Available: https://www.hindawi.com/journals/bmri/2014/598129/

[37] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Appl. Energy*, vol. 170, pp. 22–29, May 2016, doi: 10.1016/j.apenergy.2016.02.114.

[38] Ö. F. Ertugrul, "Forecasting electricity load by a novel recurrent extreme learning machines approach," *Int. J. Elect. Power Energy Syst.*, vol. 78, pp. 429–435, Jun. 2016, doi: 10.1016/j.ijepes.2015.12.006.

[39] W. Zhang, H. Hua, and J. Cao, "Short term load forecasting based on IGSA-ELM algorithm," in *Proc. IEEE Int. Conf. Energy Internet (ICEI)*, Apr. 2017, pp. 296–301, doi: 10.1109/ICEI.2017.59.

[40] J. Li, Z. Song, Y. Zhong, Z. Zhang, and J. Li, "Short-term load forecasting based on improved extreme learning machine," in *Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2017, pp. 584–588, doi: 10.1109/ICBDA.2017.8078701.

[41] Y. Chen, M. Kloft, Y. Yang, C. Li, and L. Li, "Mixed kernel based extreme learning machine for electric load forecasting," *Neurocomputing*, vol. 312, pp. 90–106, Oct. 2018, doi: 10.1016/j.neucom.2018.05.068.

[42] J. Bedi and D. Toshniwal, "Empirical mode decomposition based deep learning for electricity demand forecasting," *IEEE Access*, vol. 6, pp. 49144–49156, 2018, doi: 10.1109/ACCESS.2018.2867681.

[43] V. H. Ferreira and A. P. A. da Silva, "Toward estimating autonomous neural network-based electric load forecasters," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1554–1562, Nov. 2007, doi: 10.1109/TPWRS.2007.908438.

[44] N. Ding, C. Benoit, G. Foggia, Y. Bésanger, and F. Wurtz, "Neural network-based model design for short-term load forecast in distribution systems," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 72–81, Jan. 2016, doi: 10.1109/TPWRS.2015.2390132.

[45] P. Bokingkito, L. C. Velasco, and J. Vistal, *Week-Ahead Load Forecasting Using Multilayer Perceptron Neural Network for a Power Utility | MSU-IIT Scholarly*. Accessed: Jun. 19, 2020. [Online]. Available: https://scholarly.msuiit.edu.ph/pages/articledetails.php?artno=113&s=

[46] P.-H. Kuo and C.-J. Huang, "A high precision artificial neural networks model for short-term energy load forecasting," *Energies*, vol. 11, no. 1, p. 213, Jan. 2018, doi: 10.3390/en11010213.

[47] M. Askari and F. Keynia, "Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 5, pp. 845–852, Mar. 2020, doi: 10.1049/iet-gtd.2019.0797.

[48] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982, doi: 10.1007/BF00337288.

[49] S. Fan, C. Mao, and L. Chen, "Peak load forecasting using the self-organizing map," in *Advances in Neural Networks*. Berlin, Germany: Springer, 2005, pp. 640–647, doi: 10.1007/11427469_102.

[50] M. López, S. Valero, C. Senabre, J. Aparicio, and A. Gabaldon, "Application of SOM neural networks to short-term load forecasting: The Spanish electricity market case study," *Electr. Power Syst. Res.*, vol. 91, pp. 18–27, Oct. 2012, doi: 10.1016/j.epsr.2012.04.009.

[51] H. Li, Y. Zhu, J. Hu, and Z. Li, "A localized NARX neural network model for short-term load forecasting based upon self-organizing mapping," in *Proc. IEEE 3rd Int. Future Energy Electron. Conf. ECCE Asia (IFEEC-ECCE Asia)*, Jun. 2017, pp. 749–754, doi: 10.1109/IFEEC.2017.7992133.

[52] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, Jan. 1973, doi: 10.1080/01969727308546046.

[53] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2, pp. 191–203, Jan. 1984, doi: 10.1016/0098-3004(84)90020-7.

[54] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.

[55] Z. Xin and C. Tian-Lun, "Nonlinear time series forecast using radial basis function neural networks," *Commun. Theor. Phys.*, vol. 40, no. 2, pp. 165–168, Aug. 2003, doi: 10.1088/0253-6102/40/2/165.

[56] Y. Zhu and Y. He, "Short-term load forecasting model using fuzzy C means based radial basis function network," in *Proc. 6th Int. Conf. Intell. Syst. Design Appl.*, vol. 1, Oct. 2006, pp. 579–582, doi: 10.1109/ISDA.2006.235.

[57] Z. Yi, Z. Feng, and Z. Bingquan, "A novel ultra-short term load forecasting method based on load trend and fuzzy C-means clustering algorihm," in *Proc. Int. Conf. Power Syst. Technol.*, Oct. 2014, pp. 1–7, doi: 10.1109/POWERCON.2014.6993488.

[58] Y. Li, D. Han, and Z. Yan, "Long-term system load forecasting based on data-driven linear clustering method," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 306–316, Mar. 2018, doi: 10.1007/s40565-017-0288-x.

[59] N. Huang, W. Wang, S. Wang, J. Wang, G. Cai, and L. Zhang, "Incorporating load fluctuation in feature importance profile clustering for day-ahead aggregated residential load forecasting," *IEEE Access*, vol. 8, pp. 25198–25209, 2020, doi: 10.1109/ACCESS.2020.2971033.

[60] D. K. Ranaweera, N. F. Hubele, and G. G. Karady, "Fuzzy logic for short term load forecasting," *Int. J. Elect. Power Energy Syst.*, vol. 18, no. 4, pp. 215–222, May 1996, doi: 10.1016/0142-0615(95)00060-7.

[61] S.-J. Kang, C.-H. Woo, H.-S. Hwang, and K. B. Woo, "Evolutionary design of fuzzy rule base for nonlinear system modeling and control," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 37–45, Feb. 2000, doi: 10.1109/91.824766.

[62] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Interval type-2 fuzzy logic systems for load forecasting: A comparative study," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1274–1282, Aug. 2012, doi: 10.1109/TPWRS.2011.2181981.

[63] S. Hassan, A. Khosravi, J. Jaafar, and M. A. Khanesar, "A systematic design of interval type-2 fuzzy logic system using extreme learning machine for electricity load demand forecasting," *Int. J. Elect. Power Energy Syst.*, vol. 82, pp. 1–10, Nov. 2016, doi: 10.1016/j.ijepes.2016.03.001.

[64] D. Ali, M. Yohanna, M. I. Puwu, and B. M. Garkida, "Long-term load forecast modelling using a fuzzy logic approach," *Pacific Sci. Rev. A, Natural Sci. Eng.*, vol. 18, no. 2, pp. 123–127, Jul. 2016, doi: 10.1016/j.psra.2016.09.011.

[65] G. Cerne, D. Dovzan, and I. Skrjanc, "Short-term load forecasting by separating daily profiles and using a single fuzzy model across the entire domain," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7406–7415, Sep. 2018, doi: 10.1109/TIE.2018.2795555.

[66] M. Luy, V. Ates, N. Barisci, H. Polat, and E. Cam, "Short-term fuzzy load forecasting model using genetic–fuzzy and ant colony–fuzzy knowledge base optimization," *Appl. Sci.*, vol. 8, no. 6, p. 864, May 2018, doi: 10.3390/app8060864.

[67] K.-B. Song, Y.-S. Baek, D. H. Hong, and G. Jang, "Short-term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 96–101, Feb. 2005, doi: 10.1109/TPWRS.2004.835632.

[68] A. Azadeh, O. Seraj, and M. Saberi, "A review and comparison of fuzzy regression models for energy consumption estimation," in *Proc. 6th IEEE Int. Conf. Ind. Informat.*, Jul. 2008, pp. 1551–1555, doi: 10.1109/INDIN.2008.4618351.

[69] A. Azadeh, M. Khakestani, and M. Saberi, "A flexible fuzzy regression algorithm for forecasting oil consumption estimation," *Energy Policy*, vol. 37, no. 12, pp. 5567–5579, Dec. 2009, doi: 10.1016/j.enpol.2009.08.017.

[70] A. Azadeh, O. Seraj, and M. Saberi, "An integrated fuzzy regression—Analysis of variance algorithm for improvement of electricity consumption estimation in uncertain environments," *Int. J. Adv. Manuf. Technol.*, vol. 53, no. 5, pp. 645–660, Mar. 2011, doi: 10.1007/s00170-010-2862-5.

[71] Y.-M. Wi, S.-K. Joo, and K.-B. Song, "Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 596–603, May 2012, doi: 10.1109/TPWRS.2011.2174659.

[72] T. Hong and P. Wang, "Fuzzy interaction regression for short term load forecasting," *Fuzzy Optim. Decis. Making*, vol. 13, no. 1, pp. 91–103, Mar. 2014, doi: 10.1007/s10700-013-9166-9.

[73] T. Vantuch and M. Prilepok, "An ensemble of multi-objective optimized fuzzy regression models for short-term electric load forecasting," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–7, doi: 10.1109/SSCI.2017.8285348.

[74] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Dept. Eng. Fac. Comput. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. TR06, 2005, p. 10.

[75] V. Tereshko and A. Loengarov, "Collective decision making in honey-bee foraging dynamics," *Comput. Inf. Syst. J.*, vol. 9, no. 3, p. 1, 2005.

[76] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Oct. 2007, doi: 10.1007/s10898-007-9149-x.

[77] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: Artificial bee colony (ABC) algorithm and applications," *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 21–57, Jun. 2014, doi: 10.1007/s10462-012-9328-0.

[78] H. Safamehr and A. Rahimi-Kian, "A cost-efficient and reliable energy management of a micro-grid using intelligent demand-response program," *Energy*, vol. 91, pp. 283–293, Nov. 2015, doi: 10.1016/j.energy.2015.08.051.

[79] H. Çevik, M. Cunkas, and T. Sa, ResearchGate. (Apr. 26, 2020). *(PDF) Short-Term Load Forecasting Based on Swarm Intelligence for Smart Grids.* Accessed: Apr. 26, 2020. [Online]. Available: https://www.researchgate.net/publication/328048424_Short-Term_Load_Forecasting_Based_on_Swarm_Intelligence_for_Smart_Grids

[80] J. D. Farmer, N. H. Packard, and A. S. Perelson, "The immune system, adaptation, and machine learning," *Phys. D, Nonlinear Phenomena*, vol. 22, no. 1, pp. 187–204, Oct. 1986, doi: 10.1016/0167-2789(86)90240-X.

[81] H. Bersini and F. J. Varela, "Hints for adaptive problem solving gleaned from immune networks," in *Parallel Problem Solving from Nature*. Berlin, Germany: Springer, 1991, pp. 343–354, doi: 10.1007/BFb0029775.

[82] G. W. Rowe, *Theoretical Models in Biology: The Origin of Life, the Immune System, and the Brain*. Oxford, U.K.: Oxford Univ. Press, 1994. Accessed: Apr. 26, 2020. [Online]. Available: https://trove.nla.gov.au/version/45487798

[83] L. N. de Castro and J. Timmis, "An artificial immune network for multimodal function optimization," in *Proc. Congr. Evol. Comput. (CEC)*, vol. 1, May 2002, pp. 699–704, doi: 10.1109/CEC.2002.1007011.

[84] S. A. Hofmeyr and S. Forrest, "Architecture for an artificial immune system," *Evol. Comput.*, vol. 8, no. 4, pp. 443–473, Dec. 2000, doi: 10.1162/106365600568257.

[85] J. J. Kari, "Basic concepts of cellular automata," in *Handbook of Natural Computing*, G. Rozenberg, T. Bäck, and J. N. Kok, Eds. Berlin, Germany: Springer, 2012, pp. 3–24.

[86] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in *Proc. IEEE Comput. Soc. Symp. Res. Secur. Privacy*, May 1994, pp. 202–212, doi: 10.1109/RISP.1994.296580.

[87] J. O. Kephart, "A biologically inspired immune system for computers," in in *Proc. 4th Int. Workshop Synth. Simulation Living Syst.*, 1994, pp. 130–139.

[88] D. Dasgupta, "An overview of artificial immune systems and their applications," in *Artificial Immune Systems and their Applications*, D. Dasgupta, Ed. Berlin, Germany: Springer, 1999, pp. 3–21.

[89] Y. Yong, W. Sun'an, and S. Wanxing, "Short-term load forecasting using artificial immune network," in *Proc. Int. Conf. Power Syst. Technol.*, vol. 4, Oct. 2002, pp. 2322–2325, doi: 10.1109/ICPST.2002.1047199.

[90] G. Dudek, "Artificial immune system for short-term electric load forecasting," in *Artificial Intelligence and Soft Computing*. Berlin, Germany: Springer, 2008, pp. 1007–1017, doi: 10.1007/978-3-540-69731-2_95.

[91] J. H. Holland, *Adaptation in Atural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.

[92] W. T. Ghareeb and E. F. El Saadany, "Multi-gene genetic programming for short term load forecasting," in *Proc. 3rd Int. Conf. Electr. Power Energy Convers. Syst.*, Oct. 2013, pp. 1–5, doi: 10.1109/EPECS.2013.6713061.

[93] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1023/A:1022627411411.

[94] X.-S. Yang, "Nature-inspired mateheuristic algorithms: Success and new challenges," *J. Comput. Eng. Inf. Technol.*, vol. 1, no. 1, pp. 1–5, 2012, doi: 10.4172/2324-9307.1000e101.

[95] A. Kavousi-Fard, H. Samet, and F. Marzbani, "A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting," *Expert Syst. Appl.*, vol. 41, no. 13, pp. 6047–6056, Oct. 2014, doi: 10.1016/j.eswa.2014.03.053.

[96] L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting," *Appl. Energy*, vol. 167, pp. 135–153, Apr. 2016, doi: 10.1016/j.apenergy.2016.01.050.

[97] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," *Simulation*, vol. 76, no. 2, pp. 60–68, Feb. 2001, doi: 10.1177/003754970107600201.

[98] M. Zeng, S. Xue, Z. Wang, X. Zhu, and G. Zhang, "Short-term load forecasting of smart grid systems by combination of general regression neural network and least squares-support vector machine algorithm optimized by harmony search algorithm method," *Appl. Math. Inf. Sci.*, vol. 7, no. 1L, pp. 291–298, Feb. 2013, doi: 10.12785/amis/071L38.

[99] W.-T. Pan, "A new fruit fly optimization algorithm: Taking the financial distress model as an example," *Knowl.-Based Syst.*, vol. 26, pp. 69–74, Feb. 2012, doi: 10.1016/j.knosys.2011.07.001.

[100] H. Li, S. Guo, H. Zhao, C. Su, and B. Wang, "Annual electric load forecasting by a least squares support vector machine with a fruit fly optimization algorithm," *Energies*, vol. 5, no. 11, pp. 4430–4445, Nov. 2012, doi: 10.3390/en5114430.

[101] G. Cao and L. Wu, "Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting," *Energy*, vol. 115, pp. 734–745, Nov. 2016, doi: 10.1016/j.energy.2016.09.065.

[102] C.-H. Wu, G.-H. Tzeng, and R.-H. Lin, "A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4725–4735, Apr. 2009, doi: 10.1016/j.eswa.2008.06.046.

[103] J. Koza, "Genetic programming as a means for programming computers by natural selection," *Statist. Comput.*, vol. 4, no. 2, pp. 87–112, Jun. 1994, doi: 10.1007/BF00175355.

[104] D. G. Lee, B. W. Lee, and S. H. Chang, "Genetic programming model for long-term forecasting of electric power demand," *Electr. Power Syst. Res.*, vol. 40, no. 1, pp. 17–22, Jan. 1997, doi: 10.1016/S0378-7796(96)01125-X.

[105] C.-C. Hsu, C.-H. Wu, S.-C. Chen, and K.-L. Peng, "Dynamically optimizing parameters in support vector regression: An application of electricity load forecasting," in *Proc. 39th Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, vol. 2, Jan. 2006, p. 30c, doi: 10.1109/HICSS.2006.132.

[106] G. Zhang and J. Guo, "A novel method for hourly electricity demand forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1351–1363, Mar. 2020, doi: 10.1109/TPWRS.2019.2941277.

[107] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, Nov. 1995, pp. 1942–1948, doi: 10.1109/ICNN.1995.488968.

[108] Eberhart and Y. Shi, "Particle swarm optimization: Developments, applications and resources," in *Proc. Congr. Evol. Comput.*, vol. 1, May 2001, pp. 81–86, doi: 10.1109/CEC.2001.934374.

[109] W.-C. Hong, "Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model," *Energy Convers. Manage.*, vol. 50, no. 1, pp. 105–117, Jan. 2009, doi: 10.1016/j.enconman.2008.08.031.

[110] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3341–3350, Jul. 2018, doi: 10.1109/TSG.2016.2628061.

[111] B. Akay, "A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding," *Appl. Soft Comput.*, vol. 13, no. 6, pp. 3066–3091, Jun. 2013, doi: 10.1016/j.asoc.2012.03.072.

[112] M. A. M. Daut, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, and F. Hussin, "An improved building load forecasting method using a combined least square support vector machine and modified artificial bee colony," *Elektrika-J. Electr. Eng.*, vol. 16, no. 1, pp. 1–5, Apr. 2017. [Online]. Available: https://elektrika.utm.my/index.php/ELEKTRIKA_Journal/article/view/22, doi: 10.11113/elektrika.v16n1.22.

[113] P.-F. Pai and W.-C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Convers. Manage.*, vol. 46, no. 17, pp. 2669–2688, Oct. 2005, doi: 10.1016/j.enconman.2005.02.004.

[114] J. Wang, Y. Zhou, and X. Chen, "Electricity load forecasting based on support vector machines and simulated annealing particle swarm optimization algorithm," in *Proc. IEEE Int. Conf. Automat. Logistics*, Aug. 2007, pp. 2836–2841, doi: 10.1109/ICAL.2007.4339064.

[115] Y. Chen, W.-C. Hong, W. Shen, and N. Huang, "Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm," *Energies*, vol. 9, no. 2, p. 70, Jan. 2016, doi: 10.3390/en9020070.

[116] D. Niu and S. Dai, "A short-term load forecasting model with a modified particle swarm optimization algorithm and least squares support vector machine based on the denoising method of empirical mode decomposition and grey relational analysis," *Energies*, vol. 10, no. 3, p. 408, Mar. 2017, doi: 10.3390/en10030408.

[117] B. Wang, T. Han, and S. Zhang, "Research on short-term load forecasting based on improved support vector regression," presented at the 4th Int. Conf. Elect. Electron. Eng. Comput. Sci. (ICEEECS), Jinan, China, 2016, doi: 10.2991/iceeecs-16.2016.156.

[118] S.-M. Baek, "Mid-term load pattern forecasting with recurrent artificial neural network," *IEEE Access*, vol. 7, pp. 172830–172838, 2019, doi: 10.1109/ACCESS.2019.2957072.

[119] D. C. Park, M. A. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, May 1991, doi: 10.1109/59.76685.

[120] H.-Z. Li, S. Guo, C.-J. Li, and J.-Q. Sun, "A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm," *Knowl.-Based Syst.*, vol. 37, pp. 378–387, Jan. 2013, doi: 10.1016/j.knosys.2012.08.015.

[121] A. Kavousi-Fard, T. Niknam, and M. Golmaryami, "Short term load forecasting of distribution systems by a new hybrid modified FA-backpropagation method," *J. Intell. Fuzzy Syst.*, vol. 26, no. 1, pp. 517–522, 2014, doi: 10.3233/IFS-131025.

[122] L. Hernández, C. Baladrón, J. Aguiar, L. Calavia, B. Carro, A. Sánchez-Esguevillas, F. Pérez, Á. Fernández, and J. Lloret, "Artificial neural network for short-term load forecasting in distribution systems," *Energies*, vol. 7, no. 3, pp. 1576–1598, Mar. 2014, doi: 10.3390/en7031576.

[123] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015, doi: 10.1109/TSG.2014.2364233.

[124] F. Fahiman, S. M. Erfani, S. Rajasegarar, M. Palaniswami, and C. Leckie, "Improving load forecasting based on deep learning and K-shape clustering," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4134–4141, doi: 10.1109/IJCNN.2017.7966378.

[125] P. Ray, S. Sen, and A. K. Barisal, "Hybrid methodology for short-term load forecasting," in *Proc. IEEE Int. Conf. Power Electron., Drives Energy Syst. (PEDES)*, Dec. 2014, pp. 1–6, doi: 10.1109/PEDES.2014.7041963.

[126] C.-T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Trans. Comput.*, vol. 40, no. 12, pp. 1320–1336, Dec. 1991, doi: 10.1109/12.106218.

[127] J.-S.-R. Jang and C.-T. Sun, "Neuro-fuzzy modeling and control," *Proc. IEEE*, vol. 83, no. 3, pp. 378–406, Mar. 1995, doi: 10.1109/5.364486.

[128] A. Kaysal, S. Koroglu, Y. Oguz, and K. Kaysal, "Artificial neural networks and adaptive neuro-fuzzy inference systems approaches to forecast the electricity data for load demand, an analysis of dinar district case," in *Proc. 2nd Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2018, pp. 1–6, doi: 10.1109/ISMSIT.2018.8567270.

[129] A. Khotanzad, E. Zhou, and H. Elragal, "A neuro-fuzzy approach to short-term load forecasting in a price-sensitive environment," *IEEE Trans. Power Syst.*, vol. 17, no. 4, pp. 1273–1282, Nov. 2002, doi: 10.1109/TPWRS.2002.804999.

[130] M. B. A. Hamid and T. K. A. Rahman, "Short term load forecasting using an artificial neural network trained by artificial immune system learning algorithm," in *Proc. 12th Int. Conf. Comput. Modelling Simulation*, Cambridge, U.K., 2010, pp. 408–413, doi: 10.1109/UKSIM.2010.82.

[131] B. Jawerth and W. Sweldens, "An overview of wavelet based multiresolution analyses," *SIAM Rev.*, vol. 36, no. 3, pp. 377–412, Sep. 1994, doi: 10.1137/1036095.

[132] Z. Bashir and M. E. El-Hawary, "Short term load forecasting by using wavelet neural networks," in *Proc. Can. Conf. Electr. Comput. Eng., Navigating New Era*, Halifax, NS, Canada, vol. 1, 2000, pp. 163–166, doi: 10.1109/CCECE.2000.849691.

[133] B.-L. Zhang and Z.-Y. Dong, "An adaptive neural-wavelet model for short term load forecasting," *Electr. Power Syst. Res.*, vol. 59, no. 2, pp. 121–129, Sep. 2001, doi: 10.1016/S0378-7796(01)00138-9.

[134] C. Guan, P. B. Luh, L. D. Michel, Y. Wang, and P. B. Friedland, "Very short-term load forecasting: Wavelet neural networks with data prefiltering," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 30–41, Feb. 2013, doi: 10.1109/TPWRS.2012.2197639.

[135] B. Li, J. Zhang, Y. He, and Y. Wang, "Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test," *IEEE Access*, vol. 5, pp. 16324–16331, 2017, doi: 10.1109/ACCESS.2017.2738029.

[136] L. Alfieri and P. De Falco, "Wavelet-based decompositions in probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1367–1376, Mar. 2020, doi: 10.1109/TSG.2019.2937072.

[137] M. El-Hendawi and Z. Wang, "An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting," *Electr. Power Syst. Res.*, vol. 182, May 2020, Art. no. 106265, doi: 10.1016/j.epsr.2020.106265.

[138] M. El-Telbany and F. El-Karmi, "Short-term forecasting of jordanian electricity demand using particle swarm optimization," *Electric Power Syst. Res.*, vol. 78, no. 3, pp. 425–433, Mar. 2008, doi: 10.1016/j.epsr.2007.03.011.

[139] N. Liu, Q. Tang, J. Zhang, W. Fan, and J. Liu, "A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids," *Appl. Energy*, vol. 129, pp. 336–345, Sep. 2014, doi: 10.1016/j.apenergy.2014.05.023.

[140] C. Zhang and P. Ma, "Short-term electricity price forecasting based on PSO algorithm and RBF neural network algorithm," in *Proc. Int. Conf. Measuring Technol. Mechatronics Automat.*, vol. 3, Mar. 2010, pp. 334–337, doi: 10.1109/ICMTMA.2010.22.

[141] S. H. Ling, F. H. F. Leung, H. K. Lam, Y.-S. Lee, and P. K. S. Tam, "A novel genetic-algorithm-based neural network for short-term load forecasting," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 793–799, Aug. 2003, doi: 10.1109/TIE.2003.814869.

[142] A. Azadeh, S. F. Ghaderi, S. Tarverdian, and M. Saberi, "Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption," *Appl. Math. Comput.*, vol. 186, no. 2, pp. 1731–1741, Mar. 2007, doi: 10.1016/j.amc.2006.08.093.

[143] T. M. Ventura, A. G. De Oliveira, C. A. Martins, J. M. de Figueiredo, and R. De Souza Rosa Gomes, "Study of how the integration of artificial neural network and genetic algorithm should be made for modeling meteorological data," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 719–722, doi: 10.1109/ICMLA.2015.165.

[144] Y. K. Semero, J. Zhang, and D. Zheng, "PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy," *CSEE J. Power Energy Syst.*, vol. 4, no. 2, pp. 210–218, Jun. 2018, doi: 10.17775/CSEEJPES.2016.01920.

[145] V. S. Kodogiannis, M. Amina, and I. Petrounias, "A clustering-based fuzzy wavelet neural network model for short-term load forecasting," *Int. J. Neural Syst.*, vol. 23, no. 5, Oct. 2013, Art. no. 1350024, doi: 10.1142/S012906571350024X.

[146] J. P. Rothe, A. K. Wadhwani, and S. Wadhwani, "Artificial neural network and ANFIS based short term load forecasting in real time electrical load environment," *Int. J. Current Eng. Technol.*, vol. 4, no. 3, pp. 1939–1944, 2014.

[147] W. Sun and M. Ye, "Short-term load forecasting based on wavelet transform and least squares support vector machine optimized by fruit fly optimization algorithm," *J. Electr. Comput. Eng.*, vol. 2015, pp. 1–9, Dec. 2015, doi: 10.1155/2015/862185.

[148] Z. A. Bashir and M. E. El-Hawary, "Applying wavelets to short-term load forecasting using PSO-based neural networks," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 20–27, Feb. 2009, doi: 10.1109/TPWRS.2008.2008606.

[149] S. T. Siri, "Short-term load forecasting using a hybrid of genetic algorithm (Ga) and particle swarm optimization (Pso) for an optimized neural network," Univ. Nairobi, Nairobi, Kenya, Tech. Rep., 2018, p. 97.

[150] D. K. Chaturvedi, A. P. Sinha, and O. P. Malik, "Short load forecast using fuzzy logic and wavelet transform integrated generalized neural network," *Int. J. Elect. Power Energy Syst.*, vol. 67, pp. 230–237, May 2015, doi: 10.1016/j.ijepes.2014.11.027.

[151] X. Wang and Y. Wang, "A hybrid model of EMD and PSO-SVR for short-term load forecasting in residential quarters," *Math. Problems Eng.*, vol. 2016, pp. 1–10, Dec. 2016, doi: 10.1155/2016/9895639.

[152] Y. K. Semero, J. Zhang, and D. Zheng, "EMD-PSO-ANFIS-based hybrid approach for short-term load forecasting in microgrids," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 3, pp. 470–475, Feb. 2020, doi: 10.1049/iet-gtd.2019.0869.

[153] J. Wu, Z. Cui, Y. Chen, D. Kong, and Y.-G. Wang, "A new hybrid model to predict the electrical load in five states of australia," *Energy*, vol. 166, pp. 598–609, Jan. 2019, doi: 10.1016/j.energy.2018.10.076.

[154] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017, doi: 10.1016/j.asoc.2017.01.015.

[155] M. Jawad, S. M. Ali, B. Khan, C. A. Mehmood, U. Farid, Z. Ullah, S. Usman, A. Fayyaz, J. Jadoon, N. Tareen, A. Basit, M. A. Rustam, I. Sami, "Genetic algorithm-based non-linear auto-regressive with exogenous inputs neural network short-term and medium-term uncertainty modelling and prediction for electrical load and wind speed," *J. Eng.*, vol. 2018, no. 8, pp. 721–729, 2018. Accessed: Apr. 26, 2020. [Online]. Available: https://semanticscholar.org/paper/Genetic-algorithm-based-non-linear-auto-regressive-Jawad-Ali/81016573f8ca18dd78d14ec526a8bdd4eac1d745#paper-header

[156] Q. Liu, Y. Shen, L. Wu, J. Li, L. Zhuang, and S. Wang, "A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting," *CSEE J. Power Energy Syst.*, vol. 4, no. 2, pp. 226–237, Jun. 2018, doi: 10.17775/CSEEJPES.2016.00080.

[157] Y. K. Semero, J. Zhang, D. Zheng, and D. Wei, "An accurate very short-term electric load forecasting model with binary genetic algorithm based feature selection for microgrid applications," *Electr. Power Compon. Syst.*, vol. 46, nos. 14–15, pp. 1570–1579, Sep. 2018, doi: 10.1080/15325008.2018.1509911.

[158] P. Lv, S. Liu, W. Yu, S. Zheng, and J. Lv, "EGA-STLF: A hybrid short-term load forecasting model," *IEEE Access*, vol. 8, pp. 31742–31752, 2020, doi: 10.1109/ACCESS.2020.2973350.

[159] Y.-F. Zhang and H.-D. Chiang, "Enhanced ELITE-load: A novel CMP-SOATT methodology constructing short-term load forecasting model for industrial applications," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2325–2334, Apr. 2020, doi: 10.1109/TII.2019.2930064.

[160] W. Xu, H. Hu, and W. Yang, "Energy time series forecasting based on empirical mode decomposition and FRBF-AR model," *IEEE Access*, vol. 7, pp. 36540–36548, 2019, doi: 10.1109/ACCESS.2019.2902510.

[161] Z. Sun, S. Zhao, and J. Zhang, "Short-term wind power forecasting on multiple scales using VMD decomposition, K-means clustering and LSTM principal computing," *IEEE Access*, vol. 7, pp. 166917–166929, 2019, doi: 10.1109/ACCESS.2019.2942040.

[162] S. K. Acharya, Y.-M. Wi, and J. Lee, "Short-term load forecasting for a single household based on convolution neural networks using data augmentation," *Energies*, vol. 12, no. 18, p. 3560, Sep. 2019, doi: 10.3390/en12183560.

MD. SAMIUL HAQUE SUNNY received the B.Sc. degree in electrical and electronics engineering from the Khulna University of Engineering and Technology (KUET), in 2017. His undergraduate thesis was on High-Performance State Parameter Observation With Sensorless Vector Control of Induction Motor. He is currently working on data mining with AI algorithms, EEG signal for better BCI applications, and the structures of CNN for upgrading its performance in image recognition. His research interests include artificial intelligence, brain-computer interface, digital signal and image processing, data mining, power system stability, and robotics. His previous works were presented at the ICEEICT, in 2016, ICAEE, in 2017, IC2IT, in 2017, EICT, in 2017, and the IEEE R10-HTC.

DEBOPRIYA ROY DIPTA received the B.Sc. degree in electrical and electronics engineering from the Khulna University of Engineering and Technology, in 2018. His undergraduate thesis was on Isolated Speech Recognition System Using Convolutional Neural Network. He is currently working on the attention-based algorithm for continuous Bangla speech recognition. His research interests include machine learning, signal and image processing, computer vision, and energy systems. His previous works were presented at the IC4ME2, in 2018, ECCE, in 2019, and ICASERT, in 2019.

ABDULLAH AL MAMUN received the B.Sc. degree in electrical and electronics engineering from the Department of Electrical and Electronics Engineering, International Islamic University Chittagong (IIUC). His research interests include energy storage systems, renewable energy sources, the integration of microgrid and smart grid applications with machine learning, artificial intelligence, and robotics. He is also an active Volunteer of the IEEE and the IEEE Bangladesh Section.

MD. SOHEL received the B.Sc. degree in electrical and electronics engineering from the Department of Electrical and Electronics Engineering (EEE), International Islamic University Chittagong (IIUC). His research interests include machine learning and artificial intelligence. He has been affiliated with many IEEE related activities.

NAEEM MOHAMMAD (Student Member, IEEE) received the B.Sc. degree in electrical and electronics engineering from the Khulna University of Engineering and Technology (KUET), in 2019. His undergraduate thesis was on the IoT-Based Supervisory Control and Data Acquisition System. His research interests include smart grids, renewable energy, machine learning, data science, and the Internet of Things. He has been affiliated with various IEEE related activities.

EKLAS HOSSAIN (Senior Member, IEEE) received the B.S. degree in electrical and electronics engineering from the Khulna University of Engineering and Technology, Bangladesh, in 2006, the M.S. degree in mechatronics and robotics engineering from International Islamic University Malaysia, Malaysia, in 2010, and the Ph.D. degree from the College of Engineering and Applied Science, University of Wisconsin–Milwaukee (UWM). He was working in the areas of distributed power systems and renewable energy integration for the last ten years, and he has published a number of research articles and posters in these fields. He is now involved with several research projects on renewable energy and grid-tied microgrid system at Oregon Tech, as an Associate Professor in the Department of Electrical Engineering and Renewable Energy. He is working as an Associate Researcher with the Oregon Renewable Energy Center (OREC). He is a registered professional engineer (PE) in the state of Oregon, USA. He is also a Certified Energy Manager (CEM) and Renewable Energy Professional (REP). He, with his dedicated research team, is looking forward to explore methods to make the electric power systems more sustainable, cost-effective, and secure through extensive research and analysis on energy storage, micro-grid systems, and renewable energy sources. His research interests include modeling, analysis, design, and control of power electronic devices, energy storage systems, renewable energy sources, the integration of distributed generation systems, micro-grid and smart grid applications, robotics, and advanced control systems. He is a Senior Member of the Association of Energy Engineers (AEE). He was the Winner of the Rising Faculty Scholar Award from the Oregon Institute of Technology for his outstanding contribution to teaching. He is currently serving as an Associate Editor for IEEE Access.

• • •